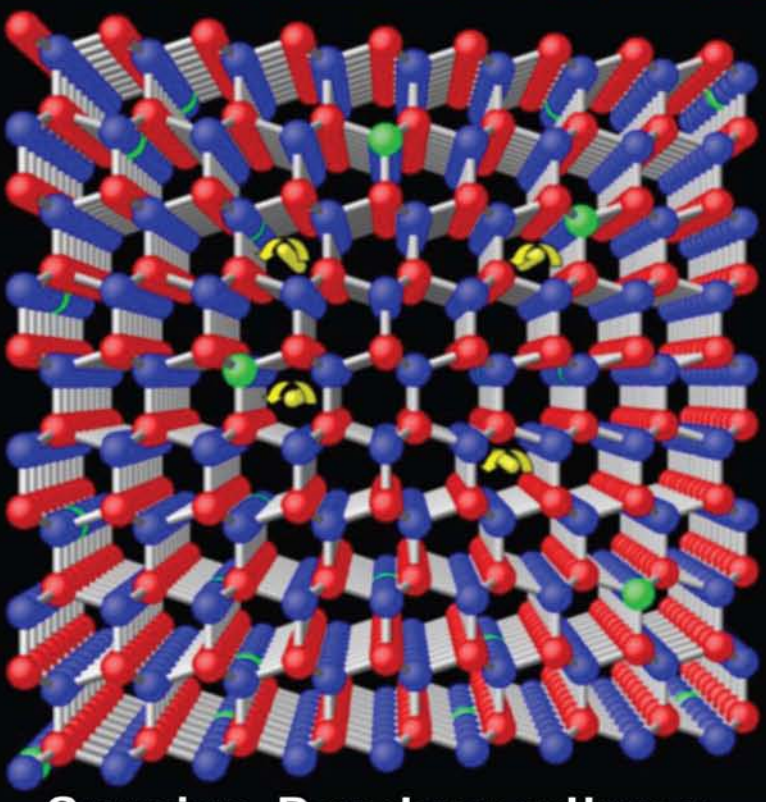
INTRODUCTION TO Spintronics



Supriyo Bandyopadhyay
Marc Cahay

CRC Press
Taylor & Francis Group

INTRODUCTION TO
Spintronics

Supriyo Bandyopadhyay
Virginia Commonwealth University
Richmond, U.S.A.

Marc Cahay
University of Cincinnati
Ohio, U.S.A.

CRC Press
Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business

CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

© 2008 by Taylor & Francis Group, LLC
CRC Press is an imprint of Taylor & Francis Group, an Informa business

No claim to original U.S. Government works
Version Date: 20130919

International Standard Book Number-13: 978-1-4200-0474-8 (eBook - PDF)

This book contains information obtained from authentic and highly regarded sources. Reasonable efforts have been made to publish reliable data and information, but the author and publisher cannot assume responsibility for the validity of all materials or the consequences of their use. The authors and publishers have attempted to trace the copyright holders of all material reproduced in this publication and apologize to copyright holders if permission to publish in this form has not been obtained. If any copyright material has not been acknowledged please write and let us know so we may rectify in any future reprint.

Except as permitted under U.S. Copyright Law, no part of this book may be reprinted, reproduced, transmitted, or utilized in any form by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying, microfilming, and recording, or in any information storage or retrieval system, without written permission from the publishers.

For permission to photocopy or use material electronically from this work, please access www.copyright.com (http://www.copyright.com/) or contact the Copyright Clearance Center, Inc. (CCC), 222 Rosewood Drive, Danvers, MA 01923, 978-750-8400. CCC is a not-for-profit organization that provides licenses and registration for a variety of users. For organizations that have been granted a photocopy license by the CCC, a separate system of payment has been arranged.

Trademark Notice: Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

Visit the Taylor & Francis Web site at
http://www.taylorandfrancis.com

and the CRC Press Web site at
http://www.crcpress.com

Dedication

S. B. dedicates this book to Bimalendu, Bela, Anuradha and Saumil Bandyopadhyay. M. C. thanks his wife Maureen for her constant support and dedicates this book to the memory of his father and father-in-law.

Preface

This is a textbook intended to introduce a student of engineering, materials science, and/or applied physics, to the field of *Spintronics*. While the term "spintronics" may have different connotations for different people, in this textbook it deals primarily with the science and technology of using the spin degree of freedom of a charge carrier to store, encode, access, process and/or transmit information in some way. That role had been traditionally delegated to the "charge" of an electron, not its "spin". Interest in spintronics is motivated by a long standing belief that replacing charge with spin in signal processing devices may yield some advantages in terms of increased processing speed, lower power consumption and/or increased device density. While this may not be always true, there are some scenarios where it *may* become true in the near future. In this textbook, we place particular emphasis on identifying situations where "spin" may have an advantage over "charge", and where it may not (see, in particular, Chapters 12-14).

The recent advent of quantum computing has added a new dimension. The spin polarization of a single electron can exist in a coherent superposition of two orthogonal spin polarizations (i.e., mutually anti-parallel spin orientations) for a relatively long time without losing the phase coherence. The charge degree of freedom, on the other hand, loses phase coherence much faster. Therefore, spin has become the preferred vehicle to host a quantum bit (or "qubit") which is a coherent superposition of two orthogonal states representing classical logic bits 0 and 1. The potential application of spin to scalable quantum logic processors has a short history, but has provided a tremendous boost to spintronics.

This textbook is expected to equip the reader with sufficient knowledge and understanding to conduct research in the field of spintronic devices, particularly semiconductor based spintronic devices. We assume that readers have first year graduate level knowledge of device engineering, solid state physics and quantum mechanics. At Virginia Commonwealth University, parts of this material were used to teach a one-semester course in spintronics to graduate students who had this background. The experiment was successful and the feedback was quite encouraging. This course is now planned to be offered in a distance learning format to graduate students in six Virginia universities, as part of an endeavor funded by a Partnership for Innovation (PFI) grant from the US National Science Foundation.

At the University of Cincinnati, parts of this material (particularly Chapters 3, 4, 5 and 14) were used to teach a course on quantum computing. That

experiment too was immensely successful and provided much of the impetus for writing this textbook.

This book is organized into fifteen chapters. The first chapter provides a historical perspective to those who have had little or no exposure to this field. It traces the early history of spin, the anomalous Zeeman effect, and ends with the accidental discovery of "spin" by Stern and Gerlach in 1922.

Chapter 2 introduces the quantum mechanics machinery needed to understand spin physics, as well as analyze spin transport and general spin dynamics in solid state structures. It also introduces the concept of Pauli spin matrices, the Pauli Equation, and finally its relativistic refinement - the Dirac Equation. Since, in this textbook, we will never encounter any situation where relativistic corrections become important, we will not have any occasion to use the Dirac Equation. The Pauli Equation will be sufficient for all scenarios. Nonetheless, it is important to gain an appreciation for the Dirac Equation, since the quantum mechanical nuances associated with spin cannot be fully absorbed without an understanding of Dirac's seminal work.

Chapter 3 introduces the Bloch sphere concept since it is a very useful tool to visualize the dynamics of a spin-1/2 particle (e.g., an electron), or qubit encoded in the spin of an electron, under the action of external magnetic fields. Applications of the Bloch sphere concept are elucidated with a number of examples. A spinor is viewed as a radial vector in the Bloch sphere, and this serves as a nice visualization tool for students interested in quantum computing and other applications of spintronics. All coherent motions of the spinor (where spin does not relax) are essentially excursions on the surface of the Bloch sphere.

Chapter 4 deals with an important application of the Bloch sphere concept, namely the derivation of Rabi oscillation and the Rabi formula for coherent spin rotation or spin flip. These have important applications in many spin-related technologies such as electron spin resonance spectroscopy, nuclear magnetic resonance, and ultimately solid state versions of quantum computing. This chapter is somewhat mathematical and 'seasons' the student to deal with the algebra (and recipes) necessary for calculating quantities that are important in spintronics. If needed, this chapter can be skipped at first and revisited later.

Chapter 5 introduces the concept of the 'density matrix', pure and mixed states, Bloch equations (that describe the temporal relaxation of spin), the Bloch ball concept, and the notion of the longitudinal (T_1) and transverse (T_2) relaxation times. Several numerical examples are also presented to strengthen key concepts. Since here we allow incoherent dynamics of the spin, the motion of the spinor is no longer constrained to the Bloch sphere. The 'Bloch sphere' actually refers only to the surface of the sphere and excludes the interior. If spin relaxes so that the norm of the sphere's radius is no longer conserved, then we have to allow excursions into the interior of the Bloch sphere. Therefore, we extend the Bloch sphere concept to the 'Bloch ball' concept. This chapter contains advanced concepts and may also be skipped at first reading.

Chapter 6 introduces the rather important topic of spin-orbit interaction which is at the heart of many spintronic devices, since it offers a "handle" to manipulate spins. We focus on the two spin-orbit interactions that are predominant in the conduction band of most semiconductors: the Rashba interaction arising from structural inversion asymmetry, and the Dresselhaus interaction arising from bulk (crystallographic) inversion asymmetry. These two interactions form the basis of spintronic field effect transistors where the current flowing between two of the transistor's terminals is modulated by influencing the spin-orbit interaction in the device via a potential applied to the third terminal. In this chapter, we also discuss an important physical phenomenon that is caused by spin-orbit interaction in a solid: the intrinsic spin Hall effect.

In Chapter 7, we derive the electron dispersion relations (energy versus wavevector) of electrons in quasi two- and one-dimensional structures (quantum wells and wires) in the presence of Rashba and Dresselhaus spin-orbit interactions, as well as an external magnetic field. We also derive the spin eigenstates, which allows us to deduce the spin polarization of carriers in any band. All this is accomplished by solving the Pauli Equation. This is an example of how the Pauli Equation is applied to solve a real life problem. The dispersion relations are important to understand many physical phenomena, a particularly intriguing example of which is the 'Spin Galvanic Effect', where spin polarization in the carrier population gives rise to an electric current (charge current) without a battery! We place special emphasis on how the dispersion relations are modified by an external magnetic field. This is important since it ultimately helps the student to appreciate how an external magnetic field can affect the performance of spin-based devices. Ultrasensitive magnetic field sensors, based on the modification of the energy dispersion relations by an external magnetic field, are a direct application of this theory. An example of this is discussed in Chapter 7.

Chapter 8 discusses spin relaxation of conduction electrons in metals and semiconductors. We focus on four primary spin relaxation mechanisms: the D'yakonov-Perel, the Elliott-Yafet, the Bir-Aronov-Pikus and hyperfine interactions with nuclear spins, since these are dominant in the conduction band of semiconductors and therefore are most important in device contexts. Because spin relaxation limits the performance of most, if not all, spin-based devices, it is a vital issue. Ultimately, the aim of all device engineers and physicists is to reduce the rate of spin relaxation in spin devices, in order to make them more robust and useful. Spin relaxation also has peculiarities that are completely unexpected and without parallel in solid state physics. We present one example where spin can relax in *time* but not in *space*.

Chapter 9 introduces the more advanced concepts of exchange and spin-spin interaction. These form the basis of ferromagnetism and also the basis of single spin computing schemes that are dealt with in Chapters 13 and 14.

Chapter 10 is an introduction to spin transport in solid state structures in the presence of spin relaxation. We focus on two basic models: the drift-

diffusion model of spin transport, and the semi-classical model that goes beyond the drift-diffusion model. The "spin" drift-diffusion model is very similar to the "charge" drift-diffusion charge model applied to bipolar transport; the "up-spin" and "down-spin" carriers assume roles analogous to electrons and holes. However, it has limitations. One limitation that we emphasize with specific examples is that it fails to describe essential features of spin transport, even qualitatively, if electrons are traveling "upstream", against the force exerted on them by an electric field. In this Chapter, we present many examples of how spin relaxes in time and space in quasi one-dimensional structures in the presence of D'yakonov-Perel' spin relaxation mechanism, since it is usually the dominant mechanism for spin relaxation in semiconductor structures. These examples are based on the semi-classical model and therefore applicable to both low field transport and high field (hot electron) transport. The semiclassical model is based on combining the Liouville equation for the time evolution of the spin density matrix with the Boltzmann Transport Equation for time evolution of the carrier momentum in the presence of scattering and external electric fields.

In Chapter 11, we discuss *passive* spintronic devices such as spin valves and devices based on the giant magnetoresistance effect. Most commercial spintronic products that are currently available (magnetic read heads for reading data in computer hard disks or entertainment systems such Apple iPods, and magnetic random access memory) utilize these passive devices. Therefore, an adequate understanding of these devices is vital for engineers. We also discuss the important notions of spin injection efficiency, spin extraction and the recently discussed spin blockade. This is a long chapter with many topics and it is intended to introduce the reader to important concepts encountered in the modern spintronics literature.

Chapter 12 introduces *active* spintronic devices, specifically what we call "hybrid spintronic devices" such as Spin Field Effect Transistors and Spin Bipolar Transistors. We explain the physical basis of how these devices operate and what their shortcomings are. We make a simple estimate of their performance figures in order to project a realistic picture of whether they are or not competitive with traditional electronic devices that are currently extant. Regardless of their actual device potential, these devices are standard bearers that aroused early interest in the field among engineers and applied physicists. They will remain forever enshrined in the 'Spintronics Hall of Fame'.

Chapter 13 introduces "monolithic spintronic devices" with Single Spin Logic paradigm as the example. This chapter goes beyond isolated devices and actually discusses complete systems, such as "all-spin-logic-gates" for combinational and sequential circuits (e.g., the arithmetic logic unit of a processor). We elucidate many pertinent issues such as power dissipation, and show that this is where "spin" may have a fundamental advantage over "charge". This is an area that has remained neglected, but deserves particular attention since, unlike quantum computing, single spin classical logic does not require phase

coherence of spin. Consequently, it is robust and much easier to implement than quantum computers.

Chapter 14 is an introduction to the field of spin-based reversible logic gates (that can, in principle, compute without dissipating energy) and spintronic embodiments of quantum computers. This is a rapidly advancing field, extremely popular among many spintronic researchers, and discoveries are made at a fast pace. This chapter is written mostly for engineers and applied physicists (not computer scientists or theoretical physicists), and should provide them with the preliminary knowledge required to delve farther into this field.

Chapter 15 is a stand-alone chapter that can be treated as an *Appendix*. At first sight, it will appear unrelated to spintronics, which it is, but it has been included for a reason. There are many instances in this book when a student will have to recollect or re-familiarize herself/himself with some key results of quantum mechanics. Rather than making a trip to the nearest library, it would be more convenient to have a 'quantum mechanics primer' handy where these key results have been re-derived. This chapter is included for completeness and comprehensiveness. The reader can refer to it if and when necessary.

This work is a preliminary attempt at writing a textbook covering some important topics in spintronics. By its very nature, it must be incomplete and omit many topics that are both important and interesting. We have focused mostly on electron spin, and with the sole exception of discussing hyperfine nuclear interactions, we have ignored nuclear spin altogether. Hence, we do not discuss such well-known phenomena as the Overhauser effect, which is more relevant to nuclear spin. Another area that we have intentionally not covered in any detail is the extremely important area of *organic spintronics*. Organic semiconductors (mostly hydrocarbons) have weak spin-orbit interactions, so that spin relaxes slowly in these materials compared to inorganic semiconductors. Hence, they have a major advantage over inorganics when it comes to applications where spin relaxation must be suppressed, such as in spin based classical or quantum computing. Some reviews have appeared in the literature covering organic spintronics. Since this field is so young and progressing so fast, we considered it premature to dwell on this topic at this time. Future editions of this textbook, if there are any, will surely address this area.

This textbook also heavily emphasizes transport phenomena as opposed to optical phenomena dealing with the interaction of polarized photons with spin-polarized electrons and holes. Delving into "opto-spintronics" would have easily added a couple hundred pages to the 500-odd pages in this textbook. Our own expertise is more in transport phenomena, which has led us to focus more on transport. However, there are many excellent books (although not "textbooks") available that deal with opto-spintronics, and the interested reader can easily find an assortment of literature in that area.

Finally, some acknowledgements are due. Many of our associates have contributed indirectly to this book. They are our students, laboratory interns and

post-doctoral research associates, past and present. They include Prof. Sandipan Pramanik of University of Alberta, Canada, who was a graduate student and then a post-doctoral researcher at Virginia Commonwealth University (VCU) at the time this book was composed, Mr. Bhargava Kanchibotla, a graduate student at VCU who provided the T_2 data in cadmium sulfide nanostructures from his experiments (figure 8.4), Mr. Sivakumar Ramanathan and Mr. Sridhar Patibandla, two graduate students at VCU working with S. B. who took the first spintronics graduate course offered at VCU by S. B. and provided valuable feedback, Mr. Harsh Agarwal, a summer undergraduate intern visiting VCU from the Banaras Hindu University Institute of Technology, Varanasi, India, who computed and generated some of the plots in Chapter 13, and Mr. Amit Trivedi, another summer undergraduate intern from the Indian Institute of Technology, Kanpur, who performed some of the calculations in Chapter 14.

At the University of Cincinnati, graduate student Mr. Junjun Wang and intern Ms. Lindsay Ficke contributed immensely to the generation of data and plots. We also thank graduate student Mr. Nishant Vepachedu for proofreading the manuscript. We remain grateful to all of them.

Finally, in spite of all our best efforts, a few typographical errors may have escaped detection. We will remain grateful to any reader who points them out to us. Our e-mails are sbandy@vcu.edu and mcahay@ececs.uc.edu.

Welcome to the world of spintronics!

TABLE 0.1
Table of Universal Constants

Free electron mass (m_0)	9.1×10^{-31} Kg.
Dielectric constant of free space (ϵ_0)	8.854×10^{-12} Farads/meter
Electronic charge (e)	1.61×10^{-19} Coulombs
Reduced Planck constant ($\hbar$)	1.05×10^{-34} Joules-sec
Bohr radius of ground state in H atom (a_0)	0.529 Å$= 5.29 \times 10^{-11}$ meters
Bohr magneton (μ_B)	9.27×10^{-24} Joules/Tesla

Contents

1 The Early History of Spin **1**
- 1.1 Spin . . . 1
- 1.2 The Bohr planetary model and space quantization . . . 3
- 1.3 The birth of "spin" . . . 4
- 1.4 The Stern-Gerlach experiment . . . 5
- 1.5 The advent of Spintronics . . . 8
- 1.6 Problems . . . 10
- 1.7 References . . . 13

2 The Quantum Mechanics of Spin **17**
- 2.1 Pauli spin matrices . . . 19
  - 2.1.1 Eigenvectors of the Pauli matrices: spinors . . . 22
- 2.2 The Pauli Equation and spinors . . . 23
- 2.3 More on the Pauli Equation . . . 25
- 2.4 Extending the Pauli Equation - the Dirac Equation . . . 26
- 2.5 The time independent Dirac Equation . . . 30
  - 2.5.1 Non-relativistic approximation to the Dirac Equation . 30
  - 2.5.2 Relationship between the non-relativistic approximation to the Dirac Equation and the Pauli Equation . . 31
- 2.6 Problems . . . 33
- 2.7 Appendix . . . 36
  - 2.7.1 Working with spin operators . . . 36
  - 2.7.2 Two useful theorems . . . 36
  - 2.7.3 Applications of the *Postulates of Quantum Mechanics* to a few spin problems . . . 38
  - 2.7.4 The Heisenberg Principle for spin components . . . 42
- 2.8 References . . . 43

3 The Bloch Sphere **45**
- 3.1 The spinor and the "qubit" . . . 45
- 3.2 The Bloch sphere concept . . . 47
  - 3.2.1 Preliminaries . . . 47
  - 3.2.2 Connection between the Bloch sphere concept and the classical interpretation of the spin of electron . . . 50
  - 3.2.3 Relationship with qubit . . . 51
  - 3.2.4 Special spinors . . . 53

3.2.5 The spin flip matrix . . . 54
3.2.6 Excursions on the Bloch sphere: the Pauli matrices revisited . . . 54
3.3 Problems . . . 58
3.4 References . . . 63

4 Evolution of a Spinor on the Bloch Sphere 65
4.1 Spin-1/2 particle in a constant magnetic field: Larmor precession . . . 65
4.1.1 Rotation on the Bloch sphere . . . 67
4.2 Preparing to derive the Rabi formula . . . 69
4.3 The Rabi formula . . . 74
4.3.1 Spin flip time . . . 77
4.4 Problems . . . 80
4.5 References . . . 88

5 The Density Matrix 89
5.1 The density matrix concept: case of a pure state . . . 89
5.2 Properties of the density matrix . . . 90
5.3 Pure versus mixed state . . . 94
5.4 Concept of the Bloch ball . . . 97
5.5 Time evolution of the density matrix: case of mixed state . . . 99
5.6 The relaxation times T_1 and T_2 and the Bloch equations . . . 103
5.7 Problems . . . 116
5.8 References . . . 128

6 Spin Orbit Interaction 129
6.1 Spin orbit interaction in a solid . . . 132
6.1.1 Rashba interaction . . . 132
6.1.2 Dresselhaus interaction . . . 134
6.2 Problems . . . 135
6.3 References . . . 137

7 Magneto-Electric Subbands in Quantum Confined Structures in the Presence of Spin-Orbit Interaction 139
7.1 Dispersion relations of spin resolved magneto-electric subbands and eigenspinors in a two-dimensional electron gas in the presence of spin-orbit interaction . . . 139
7.1.1 Magnetic field in the plane of the 2-DEG . . . 142
7.1.2 Magnetic field perpendicular to the plane of the 2-DEG . . . 150
7.2 Dispersion relations of spin resolved magneto-electric subbands and eigenspinors in a one-dimensional electron gas in the presence of spin-orbit interaction . . . 151
7.2.1 Magnetic field directed along the wire axis (x-axis) . . . 153

7.2.2 Spin components . . . 155
7.2.3 Magnetic field perpendicular to wire axis and along the electric field causing Rashba effect (i.e., along y-axis) . . . 159
7.2.4 Spin components . . . 165
7.3 Magnetic field perpendicular to wire axis and the electric field causing Rashba effect (i.e., along the z-axis) . . . 165
7.3.1 Spin components . . . 168
7.3.2 Special case . . . 168
7.4 Eigenenergies of spin resolved subbands and eigenspinors in a quantum dot in the presence of spin-orbit interaction . . . 168
7.5 Why are the dispersion relations important? . . . 174
7.6 The three types of Hall Effect . . . 174
7.6.1 Quantum spin-Hall Effect . . . 185
7.7 Problems . . . 188
7.8 References . . . 192

8 Spin Relaxation 195
8.1 The spin-independent spin-orbit magnetic field . . . 197
8.2 Spin relaxation mechanisms . . . 200
8.2.1 Elliott-Yafet mechanism . . . 200
8.2.2 D'yakonov Perel' mechanism . . . 203
8.2.3 Bir-Aronov-Pikus mechanism . . . 209
8.2.4 Hyperfine interactions with nuclear spins . . . 210
8.3 Spin relaxation in a quantum dot . . . 211
8.3.1 Longitudinal and transverse spin relaxation times in a quantum dot . . . 213
8.4 The Spin Galvanic Effect . . . 217
8.5 Problems . . . 220
8.6 References . . . 229

9 Exchange Interaction 233
9.1 Identical particles and the Pauli Exclusion Principle . . . 233
9.1.1 The Helium atom . . . 234
9.1.2 The Heitler-London model of the Hydrogen molecule . 243
9.2 Hartree and Hartree-Fock approximations . . . 246
9.3 The role of exchange in ferromagnetism . . . 248
9.3.1 The Bloch model of ferromagnetism . . . 248
9.3.2 The Heisenberg model of ferromagnetism . . . 249
9.4 The Heisenberg Hamiltonian . . . 250
9.5 Problems . . . 251
9.6 References . . . 255

10 Spin Transport in Solids 257
10.1 The drift-diffusion model . . . 257
10.1.1 Derivation of the simplified steady state spin drift-diffusion equation . . . 261
10.2 The semiclassical model . . . 264
10.2.1 Spin transport in a quantum wire: Monte Carlo simulation . . . 265
10.2.2 Monte Carlo simulation . . . 267
10.2.3 Specific examples: temporal decay of spin polarization 268
10.2.4 Specific examples: spatial decay of spin polarization . 270
10.2.5 Upstream transport . . . 270
10.3 Concluding remarks . . . 274
10.4 Problems . . . 275
10.5 References . . . 275

11 Passive Spintronic Devices and Related Concepts 277
11.1 Spin valve . . . 277
11.2 Spin injection efficiency . . . 279
11.2.1 Stoner-Wohlfarth model of a ferromagnet . . . 280
11.2.2 A simple two resistor model to understand the spin valve . . . 284
11.2.3 More advanced treatment of the spin valve . . . 287
11.2.4 A transfer matrix model . . . 294
11.2.5 Application of the Jullière formula to extract the spin diffusion length in a paramagnet from spin valve experiments . . . 309
11.2.6 Spin valve experiments . . . 309
11.3 Hysteresis in spin valve magnetoresistance . . . 310
11.4 Giant magnetoresistance . . . 314
11.4.1 Applications of the spin valve and GMR effects . . . 318
11.5 Spin accumulation . . . 322
11.6 Spin injection across a ferromagnet/metal interface . . . 327
11.7 Spin injection in a spin valve . . . 332
11.8 Spin extraction at the interface between a ferromagnet and a semiconductor . . . 338
11.9 Problems . . . 343
11.10 References . . . 346

12 Hybrid Spintronics: Active Devices Based on Spin and Charge 351
12.1 Spin-based transistors . . . 351
12.2 Spin field effect transistors (SPINFET) . . . 353
12.2.1 Sub-threshold slope . . . 359
12.2.2 The effect of non-idealities . . . 361
12.2.3 The quantum well SPINFET . . . 364

12.2.4 The SPINFET based on the Dresselhaus spin-orbit interaction . . . 365
12.3 Device performance of SPINFETs . . . 369
12.3.1 Comparison between MISFET and SPINFET . . . 370
12.3.2 Comparison between HEMT and SPINFET . . . 372
12.4 Power dissipation estimates . . . 374
12.5 Other types of SPINFETs . . . 374
12.5.1 The non-ballistic SPINFET . . . 374
12.5.2 The spin relaxation transistor . . . 377
12.6 The importance of the spin injection efficiency . . . 379
12.7 Transconductance, gain, bandwidth and isolation . . . 383
12.7.1 Silicon SPINFETs . . . 384
12.8 Spin Bipolar Junction Transistors (SBJT) . . . 385
12.9 GMR-based transistors . . . 387
12.9.1 The all-metal spin transistor . . . 387
12.9.2 The spin valve transistor . . . 390
12.10 Concluding remarks . . . 393
12.11 Problems . . . 394
12.12 References . . . 395

13 Monolithic Spintronics: All-spin logic processors 399
13.1 Monolithic spintronics . . . 399
13.1.1 Bit stability and fidelity . . . 400
13.2 Reading and writing single spin . . . 401
13.3 Single Spin Logic . . . 402
13.3.1 The universal Single Electron Logic gate: the NAND gate . . . 402
13.3.2 The input dependent ground states of the NAND gate 404
13.3.3 Ground state computing with spins . . . 412
13.4 Energy dissipation issues . . . 416
13.4.1 Energy dissipated in the gate during switching . . . 416
13.4.2 Energy dissipated in the clocking circuit . . . 421
13.5 Comparison between hybrid and monolithic spintronics . . . 421
13.6 Concluding remarks . . . 422
13.7 Problems . . . 422
13.8 References . . . 424

14 Quantum Computing with Spins 429
14.1 The quantum inverter . . . 429
14.2 Can the NAND gate be switched without dissipating energy? 434
14.3 Universal reversible gate: the Toffoli-Fredkin gate . . . 439
14.3.1 Dynamics of the T-F gate . . . 441
14.4 A-matrix . . . 442
14.5 Quantum gates . . . 442

14.5.1 The strange nature of true quantum gates: the 'square root of NOT' gate . . . 443
14.6 Qubits . . . 444
14.7 Superposition states . . . 446
14.8 Quantum parallelism . . . 447
14.9 Universal quantum gates . . . 449
14.9.1 2-qubit universal quantum gates . . . 450
14.10 A 2-qubit "spintronic" universal quantum gate . . . 450
14.10.1 The silicon quantum computer based on nuclear spins 450
14.10.2 Quantum dot-based spintronic model of universal quantum gate . . . 452
14.11 Conclusion . . . 454
14.12 Problems . . . 455
14.13 References . . . 455

15 A Quantum Mechanics Primer 459
15.1 Blackbody radiation and quantization of electromagnetic energy . . . 459
15.1.1 Blackbody radiation . . . 459
15.2 The concept of the photon . . . 460
15.3 Wave-particle duality and the De Broglie wavelength . . . 463
15.4 Postulates of quantum mechanics . . . 466
15.4.1 Interpretation of the Heisenberg Uncertainty Principle . . . 472
15.4.2 Time evolution of expectation values: the Ehrenfest Theorem . . . 475
15.5 Some elements of semiconductor physics: particular applications in nanostructures . . . 477
15.5.1 Density of states: bulk (3-D) to quantum dot (0-D) . . 477
15.6 The Rayleigh-Ritz variational procedure . . . 491
15.7 The transfer matrix formalism . . . 496
15.7.1 Linearly independent solutions of the Schrödinger equation . . . 497
15.7.2 Concept of Wronskian . . . 498
15.7.3 Concept of transfer matrix . . . 499
15.7.4 Cascading rule for transfer matrices . . . 499
15.8 Problems . . . 504
15.9 References . . . 505

1

The Early History of Spin

1.1 Spin

Most students of science and engineering know that every elementary particle, such as electrons, neutrons, photons, neutrinos, etc., has a quantum mechanical property called "spin" which can be measured (perhaps not easily, but at least in principle) and has a quantized value, including zero. The vast majority of these students mentally visualize spin as the angular momentum associated with the elementary particle spinning or rotating about its own axis (like a top or a planetary object). This mental picture, although convenient and comforting, is actually somewhat crude and certainly incomplete. Landau and Lifshitz, in their classic textbook on quantum mechanics [1], wrote "[the spin] property of elementary particles is peculiar to quantum theory.. [It] has no classical interpretation... It would be wholly meaningless to imagine the 'intrinsic' angular momentum of an elementary particle as being the result of its rotation about its own axis".

The simplistic notion of self rotation about an axis cannot explain many features of spin, such as why its magnitude cannot assume continuous values and why it is quantized to certain specific values. It also causes serious problems if taken too literally. As we will see later (Problem 1.2), if we think of an electron as a solid sphere of radius equal to the Lorentz radius $e^2/(4\pi\epsilon_0 m_0 c^2)$ (where e is the electron's charge, m_0 is the mass, c is the speed of light in vacuum, and ϵ_0 is the dielectric constant of vacuum), then the velocity on the surface of a rotating electron will have to be many times the velocity of light in vacuum if such rotation were to generate an angular momentum equal to the electron's spin. Obviously that would not be permitted by the theory of relativity. Indeed, a deep understanding of quantum mechanics is required to understand how the spin property comes about. Its origin is in relativistic quantum mechanics and really was first appreciated by Paul Andrew Maurice Dirac when he derived the Dirac Equation which is the cornerstone of relativistic quantum mechanics. Richard Feynman, noted for his distaste for mysticism in physics, wrote about the notion of "spin": "It appears to be one of the few places in physics where there is a rule which can be stated very simply, but for which no one has found a simple and easy explanation. The explanation is down deep in relativistic quantum mechanics. This probably

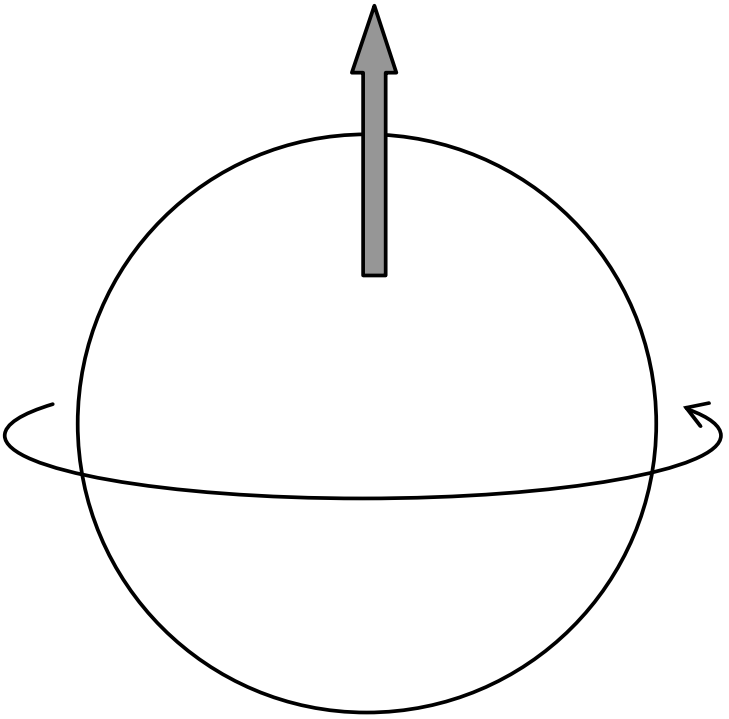


FIGURE 1.1
A mental picture to visualize spin as the angular momentum associated with self rotation of a particle about an axis. This picture, although certainly incomplete and heuristic, is adequate for many situations that the reader will encounter in this book.

means that we do not have a complete understanding of the fundamental principle involved" [2]. We do not intend to mystify "spin" any more than it already is, but rather we wish to underscore the fact that spin is *not* a classical property easily explained in terms of everyday experience. It is a property associated with relativistic quantum mechanics, for which there is no everyday experience. This is the only place in this textbook where we will mention this, since in the rest of this textbook we will have no occasion to reflect on this truism. We will be primarily involved with practical and applied aspects of spin, without relishing the fact that it is indeed an exotic property far outside what we normally deal with in the applied sciences and engineering.

1.2 The Bohr planetary model and space quantization

The history of how the concept of spin was established is somewhat tortuous. In 1913, when Niels Bohr published his theory of the hydrogen atom, he had thought of a planetary model where the electron orbits around the nucleus. The radii of the allowed orbits are quantized. Each orbit has a radius $n^2 a_0$ where a_0 is the Bohr radius of the ground state in the Hydrogen atom ($=4\pi\epsilon_0\hbar^2/m_0e^2$), which is 0.529 Å. Here $\hbar$ is the reduced Planck constant, m_0 is the electron's mass, ϵ_0 is the permittivity (dielectric constant) of free space, and e is the magnitude of its charge. The quantity n is an integer called the "principal quantum number", and it takes positive non-zero values 1, 2, 3, etc. Later Arnold Sommerfeld (in 1916), and independently Peter Debye, introduced two more quantum numbers l and m which were called the "orbital" and "magnetic" quantum numbers. While the principal quantum number determines the radius (or size) of the orbit, the angular quantum number l determines its shape. It also determines the angular momentum associated with the orbital motion in units of $\hbar$. The integers n and l obey the relation $n \geq l$. If the atom is placed in a magnetic field, the component of its angular momentum along the field takes on quantized values of $m\hbar$. The number m is an integer and satisfies the inequality $-l \leq m \leq l$. This last inequality limits the number of m values to $2l + 1$, and accordingly, the number of allowed directions of the angular momentum vector in a magnetic field is $2l + 1$. This is known as *space quantization of angular momentum.*

The energy of an electron in an atom was thought to be determined by the three quantum numbers n, l and m. When an electron makes a transition from one energy state to another, the transition involves a change in one or more of these quantum numbers. In the process of transition, the electron absorbs or emits light of a particular frequency ν which is determined by the energy difference between the initial and final states in accordance with conservation of energy:

$$E_{final} - E_{initial} = 2\pi\hbar\nu = h\nu, \tag{1.1}$$

where $E_{initial}$ and E_{final} are the electron's energy in the initial and final states, respectively.

However, when an atom is placed in a magnetic field and the spectra of emitted and absorbed light are measured, it is found that the multiplicity of the spectra (meaning all the observed frequencies) cannot be explained by the space quantization rules (that means allowed values of n, l and m) alone. In 1920, Sommerfeld tried to explain the multiplicity by invoking yet another quantum number j that he called the *inner* quantum number. However, this was not able to completely explain multiplicity. Additional frequencies (where each line split into two) were observed in a strong magnetic field and

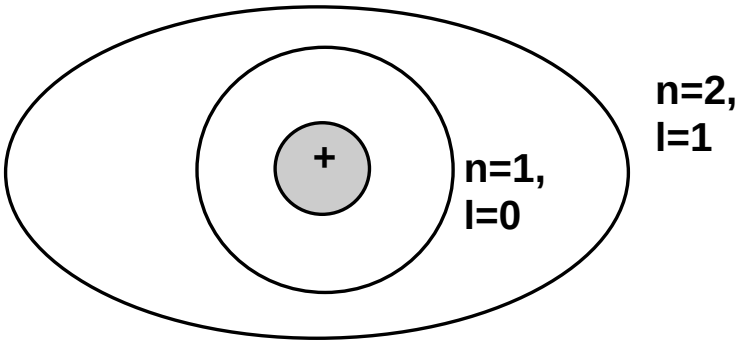


FIGURE 1.2
The Bohr planetary model of electron orbits around the nucleus of an atom (not to scale).

was referred to as the anomalous Zeeman effect. This effect could not be explained by space quantization.

1.3 The birth of "spin"

In 1925, the anomalous Zeeman effect led a young American scientist Ralph de Laer Kronig to postulate that in addition to the orbital angular momentum, an electron has an additional angular momentum caused by spinning about its own axis. Kronig postulated that the angular momentum associated with this self-rotation has a fixed magnitude of $(1/2)\hbar$. With this additional angular momentum, he was able to explain the multiplicity of spectra, and presented his idea to Wolfgang Pauli who was not impressed by it (Pauli had his own ideas to explain the multiplicity of spectra, which later turned out to be wrong). Kronig himself was not very confident because his calculations based on this spinning electron model still did not completely explain every feature of the experimentally observed spectra. More importantly, the idea of a self rotating electron presents a conundrum within the framework of classical theory. If one thinks of the electron as a sphere of radius $r_e = e^2/(4\pi\epsilon_0 mc^2)$, as

considered by Lorentz, then the rotation rate required to produce an angular momentum of $(1/2)\hbar$ is so high that the electron's surface reaches a speed more than 60 times that of light in vacuum, in stark violation of Einstein's theory of relativity! Because of this apparent contradiction, Kronig never published his ideas. Six months later, Uhlenbeck and Goudsmit published essentially the same spinning electron idea that Kronig had come up with in the journal *Naturwissenschaften*. Actually, when they realized the problem with the surface speed (see Problem 1.2), they tried to hurriedly withdraw their paper, but it was too late. The paper appeared in print. Kronig sent a letter to the British journal *Nature* criticizing the idea of Uhlenbeck and Goudsmit pointing out the problems with the spinning electron model. Meanwhile, in a second paper that appeared in *Nature*, Uhlenbeck and Goudsmit pointed out that their theory did not quite explain the experimental observations of atomic spectra. In fact there was a discrepancy by a factor of 2 with the experimental results. Later L. H. Thomas showed that this discrepancy comes about because of an incorrect definition of the electron rest system. All is well if the electron is moving at a constant velocity, but in reality it is not moving with a constant velocity because the direction of the velocity is always changing in a closed orbit. Therefore, the electron is always accelerating or decelerating. As a result, the electron's energy states are modified, and this leads to the 'factor of 2 correction'. Thomas published his findings in a letter to *Nature* in February 1926, following which, all discrepancies between theory and experiment could be resolved. This made a convert out of Pauli who ultimately endorsed the spinning electron idea of Kronig, Uhlenbeck and Goudsmit (KUG).

The spinning electron picture remains in vogue although it obviously raises unanswered questions such as why the spin angular momentum is quantized, or why the idea cannot be reconciled with the postulates of the special theory of relativity. These unanswered questions merely reveal the inadequacy of the model. The reader should be aware of this inadequacy, but that should not affect her/his understanding of any of the topics discussed in this book.

1.4 The Stern-Gerlach experiment

At least three years before the KUG ideas were published, the spin of an electron was already *unwittingly* measured in the famous Stern-Gerlach experiment, which remains a watershed event in the history of spin. Otto Stern and Walther Gerlach in Frankfurt, Germany took the space quantization business literally and designed an experiment to verify it. To them, the space quantization business concerned only the quantum numbers m, l and n, but no spin, since the spin idea had not yet been thought of by KUG. Stern and others

had thought that a gas of Hydrogen atoms will be magnetically birefringent, because the electron in the atom would orbit in a plane perpendicular to the magnetic field. This, they thought, would limit the electron's orbital angular momentum to $\pm\hbar$. It is not clear why they thought this, but Stern was convinced that an atomic beam of Hydrogen would split into two beams in a magnetic field, and despite the smearing effect of the velocity distribution, they should split far apart in a strong field that the oppositely directed components would be deflected outside the width of the original beam. If that happens, then one should observe *two* distinct beams and the birefringence will be demonstrated unequivocally. This would have been a remarkable experiment anyway, since classical theory predicted that the atomic magnets would precess in the field but remain randomly oriented so that the beam would be broadened, but not split into two. Thus, any birefringence would demonstrate that quantum physics supersedes classical physics, and it would be a *literal* demonstration of space quantization which is a quantum mechanical phenomenon outside the realm of classical physics. Bohr himself was not so convinced of his own model of the atom and did not think one could take quantum physics so literally. He thought that space quantization was a symbolic expression and convenient for calculating atomic spectra, but could not be taken literally. However, Stern was fortunately not dissuaded by Bohr and found a convert in Gerlach, who, until that time, had apparently not heard of space quantization!

The actual Stern-Gerlach experiment took place a year after it was conceived by Stern. It was not Hydrogen atoms, but a beam of Silver atoms produced by the effusion of the metal from an oven heated to 1000°C, that was collimated by two narrow slits 0.03 mm wide. It traversed a magnet 3.5 cm long that produced a magnetic flux density of 0.1 Tesla and a field gradient of 10 Tesla/cm. The calculated splitting of the beam was only 0.2 mm so that extreme alignment accuracy was required. When the experiment was completed, Stern and Gerlach could not even see the traces of the silver beam on the collector plate, so that it was impossible to determine if the splitting was there. In an episode that has now become famous, Stern's breath on the plate made history. He used to smoke cigars and the sulfur from his breath turned the silver into silver sulfide, which is jet black and is easily visible. The results of this experiment however were still inconclusive and would not have convinced a sceptic. After many efforts, Stern and Gerlach met in Göttingen in 1922 and decided to give up. But a railroad strike detained Gerlach and he attempted the experiment once again with improved accuracy of alignment. Finally a clear signature of beam splitting was observed.

We now know that the Stern-Gerlach experiment had nothing to do with space quantization, contrary to what they believed. In fact, Gerlach could never reproduce the experiment with Sodium atoms instead of Silver atoms, and Einstein questioned their interpretation (correctly). The agreement of the experiment with Stern's and Gerlach's picture of space quantization (or the Bohr model) was nothing but a coincidence. The orbital angular momentum

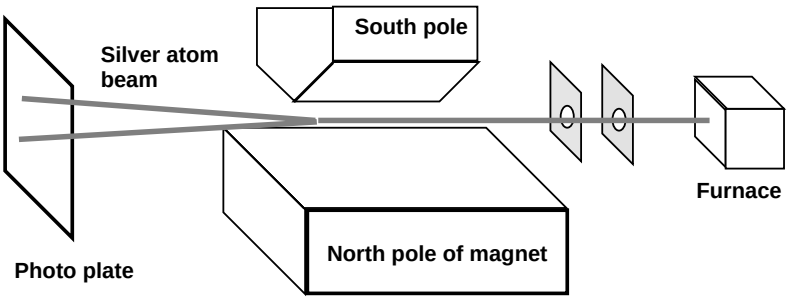


FIGURE 1.3
Experimental arrangement of the Stern-Gerlach apparatus (not to scale).

of Silver atom is actually zero, not $\hbar$, contrary to what was presumed by Stern and Gerlach. The magnetic moment of Silver atoms is therefore due solely to the *spin angular momentum* (and has nothing to do with the orbital angular momentum), which accounts for the observed splitting of the beam into two. Stern and Gerlach had *unknowingly measured the spin angular momentum.* This is the first convincing experimental observation of "spin".

It is surprising that even though the Stern-Gerlach experiment was reported in 1922 and was widely known among physicists, the postulation of the electron spin in 1925 did not immediately lead to a re-interpretation of the experiment as being a demonstration of the spin. The earliest attribution of the splitting to spin was reported in 1927 when Ronald Fraser noted that the ground state orbital angular momentum and associated magnetic moments of Silver, Hydrogen and Sodium atoms are zero. Therefore, obviously the splitting could not be due to orbital angular momentum; it had to be due to "spin".

In February of 2002, a team of physicists and chemists tried to re-enact the Stern-Gerlach experiment with sulfur tainted breath and all [3]. For a hilarious account of this re-enactment, see the December 2003 issue of the magazine *Physics Today.* Mere sulfurous breath was not sufficient to reveal traces of silver on a detecting plate, but direct exposure to cigar smoke was enough to reveal the traces.

1.5 The advent of Spintronics

Although "spin" plays a fundamental role in explaining the multiplicity of atomic spectra, this is not its most important role. It was realized in the middle of the 20th century that spin plays a fundamental role in magnetism. Every theoretical model postulated to explain the physical origin of magnetism invoked "spin" in some way or the other. This included the Bloch model, the Heisenberg model, the Stoner model and every other model advanced. While magnetism remains the domain of spin, in the late 20th century, it was realized that spin, alone or in conjunction with charge, can be harnessed to process information, particularly digital information encoded with binary bits 0 and 1. This is the central theme in the field of *Spintronics* today (see, for example, refs [4, 5]). Of course, spintronics is much more than just information rendition and includes the more traditional areas of *magnetoelectronics* that deals with magnetic or magneto-resistive effects for sensing and storing information. Early successes in this area include the developments of read heads for sensing massively dense magnetic storage media (these read heads are now routinely used in laptop computers and entertainment systems such as Apple iPods), non-volatile magnetic random access memory (MRAM) [6], programmable spintronic logic devices based on magnetic tunnel junction elements [7], rotational speed control systems [8], positioning control devices in robotics and related systems (such as automobile braking systems) [9], perimeter defense systems, magnetometers and high current monitoring devices for power systems [10], etc. Many of these developments were fueled by the investigation of how spin-polarized electric currents can be injected into ferromagnetic/paramagnetic multilayers which, in the 1980s, had led to the important discovery of the phenomenon of giant magnetoresistance (GMR) [11, 12]. Several books and review articles have been written on magnetoelectronics [13, 14, 15, 16, 17, 18, 19]. The reader is referred to these books to gain an understanding of magnetoelectronics. In this book, we will focus more on the "information technology" (IT) applications of spin which is the basis of modern spintronics.

The application of spintronics in information technology, particularly for computing and signal processing is a relatively new field. Early efforts in this area were concerned with developing spin-based analogs of conventional signal processing devices such as transistors - both the field effect type [20, 21] and the bipolar junction type [22, 23]. In this book, we call these devices *hybrid spintronic devices*, since information is still encoded in the "charge" (voltage across or current through the device), but "spin" augments the functionality of the device and *may* improve device performance. In Spin Field Effect Transistors (SPINFETs) or Spin Bipolar Junction Transistors (SBJTs), information is still processed by modulating the charge current flowing between two terminals via the application of either a voltage or a current at the third

terminal. However, the process by which the voltage or current at the third terminal exercises control over device behavior is spin-mediated. Therefore "spin" plays somewhat of a secondary role in these devices, while "charge" retains the primary role. These devices are interesting and demonstrate how spin can play a role in information handling. They are discussed in Chapter 12.

The more radical branch of spintronics is what we call *monolithic spintronics*. Here charge has no direct role and information is encoded entirely in the spin polarization of a single electron, which is made to have only two values: "up-spin" and "down-spin", by placing the electron in a magnetic field. The down-spin orientation could correspond to polarization parallel to the magnetic field and the up-spin orientation is anti-parallel to the magnetic field. These two values could represent binary bits 0 and 1. Boolean logic circuits can be fashioned from interacting single electron spins by properly engineering the interactions. An embodiment of this approach is the so-called "single spin logic" (SSL) idea [26, 27] discussed in Chapter 13. What makes this idea attractive is that bits can be flipped by simply toggling an electron's spin, without physically moving it in space and causing a current flow. This results in much reduced heat dissipation in the circuits.

The ultimate rendition of spintronic computing circuits is spin-based quantum computers, which dissipate no energy at all to complete a logic operation, since they operate on the basis of *reversible* quantum dynamics. Recently, there has been a great deal of interest to encode a quantum bit (*qubit*) using the spin degree of freedom of a single electron confined in a quantum dot [28, 29, 30, 31] or bound to a donor atom [32] to implement a quantum logic gate. The quantum mechanical phase coherence of "spin" is much longer lived than that of "charge"; consequently, spin is a natural choice for building solid state scalable quantum logic processors. Spin-based quantum computing is a rapidly expanding field of research endeavor and Chapter 14 in this textbook discusses some of the basic ideas in this field. Unfortunately, Chapter 14 will be almost certainly outdated very quickly, perhaps even by the time this book appears in print, since extremely rapid strides are being made in this field. Suffice it to say then that Spintronics is now poised at a critical juncture where technological breakthroughs may be just around the corner. Therefore, understanding the science and technology of spintronics has become imperative for students of electrical engineering, physics, and materials science.

1.6 Problems

- **Problem 1.1**

  Using the Bohr model of the Hydrogen atom, show that the magnetic moment associated with an electron moving in the lowest circular orbit is the Bohr magneton given by

$$\mu_B = \frac{e\hbar}{2m_0} \ . \tag{1.2}$$

  Show also that the magnetic flux density B_n associated with the n-th orbital is given by

$$\mu_B B_n = \frac{1}{4}(\frac{\alpha^4 Z^4}{n^5})m_o c^2 \ , \tag{1.3}$$

  where m_o is the rest mass of the electron, c is the speed of light in vacuum, Z is the atomic number of the atom, n is the principal quantum number and α is the fine structure constant.

  Calculate B_n in Tesla for the electron in the first orbit of the hydrogen atom, i.e., Z=1, n=1. Use $m_o c^2 \approx 0.5 MeV$.

  Solution

  Consider a nucleus with Z protons. The total (kinetic + potential) energy of an electron moving around this nucleus is conserved and is given by:

$$E = \frac{1}{2}m_o v^2 - \frac{Ze^2}{4\pi\epsilon_o r} \ , \tag{1.4}$$

  where e is the magnitude of the charge of the electron, v is the electron's orbital velocity and ϵ_o is the dielectric constant of vacuum. The first term in the right hand side is the kinetic energy of the electron and the second term is the (electrostatic) potential energy. We ignore any gravitational potential energy since it is negligible.

  According to Newton's Law, the centripetal force should be equal to Coulomb force; hence, we have:

$$m_o \frac{v^2}{r} = \frac{Ze^2}{4\pi\epsilon_o r^2} \ . \tag{1.5}$$

  Using Bohr's quantization condition for the angular momentum around the n-th orbit:

$$L_n = m_o v_n r = n\hbar \ , \tag{1.6}$$

where the subscript n is the principal quantum number.

We get from Equations (1.5) and (1.6):

$$L_n v_n = \frac{Ze^2}{4\pi\epsilon_o} \ . \tag{1.7}$$

This yields

$$v_n = \frac{Ze^2}{4\pi\epsilon_o \hbar}\frac{1}{n} \ . \tag{1.8}$$

We can compare this velocity to that of light and get:

$$\frac{v_n}{c} = \frac{Z}{n}\frac{e^2}{4\pi\epsilon_o \hbar c} \ , \tag{1.9}$$

where the constant $\frac{e^2}{4\pi\epsilon_o \hbar c} = \alpha$ and is referred to as the fine structure constant.

Equation (1.9) shows that an electron in the first orbit (ground state) of the hydrogen atom ($Z = n = 1$) has a velocity equal to $\alpha \approx \frac{1}{137}$ times the speed of light.

The radius of the various orbits allowed by Bohr's space quantization condition can then be obtained from Equation (1.6)

$$m_o v_n r_n = n\hbar \ . \tag{1.10}$$

We get

$$r_n = (\frac{4\pi\epsilon_o}{Ze^2})(\frac{n^2\hbar^2}{m_o}) \ . \tag{1.11}$$

This is also written as: $r_n = n^2 a_Z$, where a_Z is the effective Bohr radius given by

$$a_Z = \frac{4\pi\epsilon_o \hbar^2}{m_o Ze^2} \ . \tag{1.12}$$

The latter is equal to 0.529 $\mathring{A}$ for the Hydrogen atom ($Z = 1$).

The total energy of an electron in the n-th orbit is found from Equation (1.4) and (1.8) to be:

$$E_n = \frac{1}{2}m_o {v_n}^2 - \frac{Ze^2}{4\pi\epsilon_o r_n} = -\frac{m_o e^4 Z^2}{32{\epsilon_o}^2\hbar^2} \ . \tag{1.13}$$

This is the *binding energy* of the electron. It is also called the *ionization energy*, since this is the minimum amount of energy that will be required to liberate the electron from the nuclear attraction and make it free, thereby ionizing the parent atom. This ionization energy is 13.6 eV for the ground state of the hydrogen atom.

According to Biot-Savart's law, an electron moving around the nucleus in one of these orbits will feel a magnetic field given by:

$$\vec{B} = \frac{\vec{\mathcal{E}} \times \vec{v}}{2c^2} \ , \tag{1.14}$$

where $\mathcal{E}$ is the electric field experienced by the moving electron and the factor 2 is due to Thomas's correction for an accelerating or decelerating electron in a closed orbit.

From Coulomb's law, the electric field in the n-th orbital is given by

$$\mathcal{E}_n = \frac{Ze}{4\pi\epsilon_o {r_n}^2} \ . \tag{1.15}$$

Using Equations (1.8), (1.14) and (1.15), the magnitude of the magnetic field in the n-th orbit is

$$|\vec{B_n}| = \frac{\mathcal{E}_n v_n}{2c^2} = \frac{1}{2c^2}(\frac{Ze}{4\pi\epsilon_o {r_n}^2})v_n \tag{1.16}$$

and it is directed perpendicular to the orbital plane of the electron.

Finally, using Equations (1.11), (1.8), (1.16) and the expression for Bohr magneton given, one finds that

$$\mu_B B_n = \frac{1}{4}(\frac{\alpha^4 Z^4}{n^5})m_o c^2 \ . \tag{1.17}$$

From the above, the magnetic flux density associated with an electron orbiting the Hydrogen atom in the first orbit is 6.16 Tesla.

- **Problem 1.2**

  Show that in the classical spinning electron model, the electron's surface speed must be more than 60 times the speed of light in order to produce an angular momentum of $(1/2)\hbar$. This is why, it is inappropriate to think in classical terms that the spin of an electron is associated with rotation about its own axis.

  Solution

  The angular momentum is

$$m_0 v_s r_e = (1/2)\hbar \ , \tag{1.18}$$

where v_s is the speed on the surface of the electron and r_e is the Lorentz radius of the electron. Solving the above equation with universal constants $m_0 = 9.1\times10^{-31}$ Kg, $\hbar = 1.05\times10^{-34}$ Joules-sec, and $r_e = e^2/(4\pi\epsilon_0 m_0 c^2) = 2.8$ femtometers,

$$v_s \approx 67c. \tag{1.19}$$

Therefore, the speed of rotation on the surface of the electron is more than 60 times the speed of light.

What does this apparent fallacy imply? It tells us that the concept of spin is inherently quantum mechanical and cannot be described within the framework of classical mechanics as done here. Furthermore, the electron cannot be visualized as a nearly point charge with the Lorentz radius.

1.7 References

[1] L. D. Landau and E. M. Lifshitz, *Quantum Mechanics (Non-relativistic Theory)*. 3rd. edition (Pergamon, New York), p. 198.

[2] R. P. Feynman, R. B. Leighton and M. Sands, *The Feynman Lectures on Physics*, Vol. 3, (Addison-Wesley, Reading, MA, 1965), p. 4-3.

[3] B. Friedrich and D. Hershbach, "Stern and Gerlach: How a Bad Cigar Helped Reorient Atomic Physics", *Physics Today*, **56**, pp. 53-59 (December 2003).

[4] D. D. Awschalom, M. E. Flatte and N. Samarth, "Spintronics", Scientific American, **286**, 66, June 2002.

[5] S.A. Wolf and D.M. Treger, "Special issue on spintronics", Proc. IEEE, **91**, 647 (2003).

[6] P.P. Freitas, F. Silva, N.J. Oliveira, L.V. Melo, L. Costa and N. Almeida, "Spin valve sensors", Sensors and Actuators A, **81**, 2 (2000).

[7] J. Wang, H. Meng and J.P. Wang, "Programmable spintronics logic device based on a magnetic tunnel junction element", J. Appl. Phys., **97**, 100509 (2005).

[8] P.P. Freitas, L. Costa, N. Almeida, L.V. Melo, F. Silva, J. Bernardo and C. Santos, "Giant magnetoresistive sensors for rotational speed control", J. Appl. Phys., **85**, 5459 (1999).

[9] W.J. Ku, P.P. Freitas, P. Compadrinho and J. Barata, "Precision X-Y robotic object handling using a dual GMR bridge sensor", IEEE Trans. Magn., **36**, 2782 (2000).

[10] J. Pelegri, J.B. Egea, D. Ramirez and P.P. Freitas, "Design, fabrication, and analysis of a spin-valve base current sensor", Sensors and Actuators A, **105**, 132 (2003).

[11] M.N. Baibich, J. M. Broto, A. Fert, F. Nguyen Van Dau, F. Petroff, P. Eitenne, G. Creuzet, A. Friederich and J. Chazelas, "Giant Magnetoresistance of (001)Fe/(001)Cr Magnetic Superlattices", Phys. Rev. Lett., **61**, 2472 (1988).

[12] G. Binasch, P. Grünberg, F. Saurenbach and W. Zinn, "Enhanced magnetoresistance in layered magnetic structures with antiferromagnetic interlayer exchange", Phys. Rev. B, **39**, 4828 (1989).

[13] A detailed review of the work on giant magnetoresistance using magnetic multilayers is given by M.A.M. Gijs and G.E.W. Bauer, "Perpendicular giant magnetoresistance of magnetic multilayers", Adv. Phys., **46**, 285 (1997).

[14] S.S.P. Parkin, X. Jiang, C. Kaiser, A. Panchula, K. Roche and M. Samant, "Magnetically engineered spintronic sensors and memory", Proc. IEEE, **91**, 661 (2003).

[15] D.D. Awschalom and J.M. Kikkawa, "Electron spin and optical coherence in semiconductors", Physics Today, **52**, 33 (1999).

[16] *Magnetic Multilayers and Giant Magnetoresistance*, Ed. U. Hartman (Springer, Berlin, 2000).

[17] *Spin Electronics*, Eds. M. Ziese and M. J. Thornton, (Springer, New York, 2001).

[18] *Spin Dependent Transport in Magnetic Nanostructures*, Eds. S. Maekawa and T. Shinjo, (Taylor and Francis, New York, 2002).

[19] *Magnetic Interactions and Spin Transport*, Eds. A. Chtchelkanova, S. Wolf and Y. Idzerda, (Kluwer Academic Press, Dordrecht, 2003).

[20] S. Datta and B. Das, "Electronic analog of the electro-optic modulator", Appl. Phys. Lett., **56**, 665 (1990).

[21] S. Bandyopadhyay and M. Cahay, "Alternate spintronic analog of the electro-optic modulator", Appl. Phys. Lett., **85**, 1814 (2004).

[22] J. Fabian, I Zutic and S. Das Sarma, "Magnetic bipolar transistors", Appl. Phys. Lett., **84**, 85 (2004).

[23] M. E. Flatté and G. Vignale, "Unipolar spin diodes and transistors", Applied Physics Letters, Vol. 78, 1273 (2001); M. E. Flatté, Z. G. Yu,

E. Johnston-Halperin and D. D. Awschalom, "Theory of semiconductor magnetic bipolar transistors", Appl. Phys. Lett., **82**, 4740 (2003); M. E. Flatté and G. Vignale, "Heterostructure unipolar spin transistors", J. Appl. Phys., **97**, 104508 (2005).

[24] S. Bandyopadhyay and M. Cahay, "Re-examination of some spintronic field effect device concepts", Appl. Phys. Lett., **85**, 1433 (2004).

[25] S. Bandyopadhyay and M. Cahay, "Are spin junction transistors useful for signal processing?", Appl. Phys. Lett., **86**, 133502 (2005).

[26] S. Bandyopadhyay, B. Das and A. E. Miller, "Supercomputing with spin polarized single electrons in a quantum coupled architecture", Nanotechnology, **5**, 113 (1994).

[27] S. Bandyopadhyay, "Computing with spins: from classical to quantum computing", Superlat. Microstruct., **37**, 77 (2005).

[28] S. Bandyopadhyay and V. P. Roychowdhury, "Switching in a reversible spin logic gate", Superlat. Microstruct., **22**, 411 (1997).

[29] A. M. Bychkov, L. A. Openov and I. A. Semenihin, "Single electron computing without dissipation", JETP Lett., **66**, 298 (1997).

[30] D. Loss and D.P. DiVincenzo, "Quantum computation with quantum dots", Phys. Rev. A, **57**, 120 (1998).

[31] S. Bandyopadhyay, "Self assembled nanoelectronic quantum computer based on the Rashba effect in quantum dots", Phys. Rev. B, **61**, 13813 (2000).

[32] B.E.Kane, "A silicon-based nuclear spin quantum computer", Nature (London), **393**, 133 (1998).

2

The Quantum Mechanics of Spin

In the 1920s, the old quantum theory was gradually being superseded by the new quantum theory. The cornerstone of the old theory was Bohr's model of the hydrogen atom which predicted that an electron cannot orbit the proton in the hydrogen atom in any arbitrary fashion. Orbits are "quantized", meaning that only certain sizes, shapes and magnetic properties are allowed. The principal quantum number n determined the allowed radii of the orbits, the orbital quantum number l determined the allowed shapes, and the magnetic quantum number m determined the magnetic behavior. Additionally, there is a fourth quantum number s which denotes the fact that the electron has an additional angular momentum, loosely associated with self rotation about its own axis, and that is quantized in units of $(1/2)\hbar$. The old quantum theory was useful to infer the existence of discrete energy levels in atoms, calculate energy spacings between these levels, and therefore allowed one to interpret atomic spectra.

The new quantum theory appeared to be more revolutionary and more powerful. It was triggered by Heisenberg's discovery of *matrix mechanics* and Schrödinger's discovery of *wave mechanics*. These two formalisms would not only predict the quantization of energy and provide a prescription to determine the energy difference between the levels (and thus explain the multiplicity of atomic spectra), but also allow one to calculate easily *probabilities of transitions* between different quantized energy states. At first, matrix mechanics and wave mechanics looked entirely different in their mathematical appearance and physical meaning. However, Schrödinger and Eckart [1] independently showed that the two theories are mathematically equivalent. Towards the end of 1926, Dirac unified the two theories using the concept of state vector and thus established the *transformation theory* of quantum mechanics. This ultimately had a profound implication for the quantum mechanical (mathematical) recipe to treat spin, as we will show in this chapter.

The transformation theory is the mathematical recipe to handle modern quantum mechanics. In Heisenberg's matrix mechanics, a physical quantity is expressed by a matrix, whereas in Schrödinger's wave mechanics, a physical quantity is expressed by a linear operator. In the unified transformation theory, physical quantities are represented by abstract linear operators called Dirac's *q-numbers* which are linear operators in an infinite-dimensional linear space. Depending upon which types of orthogonal coordinate systems are used in this linear space, either matrix mechanics or wave mechanics emerges.

In other words, by using coordinate transformation in this linear space, we can derive matrix mechanics from wave mechanics and vice versa. Therefore, this unified theory was named transformation theory. The state of a quantum mechanical object is represented by a so-called *state vector* which is an abstract vector in this linear space (the "wavefunction" in wave mechanics is an example of this) and the linear space is called the *state space*.

Earlier D. Hilbert and J. von Neumann had introduced the notion of a linear space that could absorb the mathematics of matrices and vectors, as well as the mathematics of linear operators and functions. This so-called *Hilbert space* admitted a finite or denumerably (countably) infinite number of co-ordinate axes. Therefore, a state vector in the Hilbert space could have at most a denumerably infinite number of mutually orthogonal components. Dirac extended this concept to a non-denumerably infinite number of coordinate axes in his linear space via the introduction of his famous δ-function [3]. The state vector therefore could have non-denumerably infinite number of mutually orthogonal components and could be expressed as

$$\psi(q), \quad q \in [q_1, q_2] \ , \tag{2.1}$$

where the variable q is a continuous variable in the domain $[q_1, q_2]$. On the other hand, if the coordinate axes are countable, then the state vector would be expressed as

$$\psi_n, \quad n = 1, 2, 3, ... \tag{2.2}$$

where the variable n is an integer.

According to Dirac's transformation theory, the state vector (i) evolves in time according to a unitary transformation, and (ii) satisfies a *first order* differential equation with respect to time. This second property is very important as we shall see later. Depending on whether the physical quantity* represented by the state vector will yield discrete or continuous values upon measurement, the eigenvalues of the linear operator describing this physical quantity will have discrete or continuous values. Accordingly, the coordinate axes in the linear space will be discrete or continuous, and the state vector will be ψ_n or $\psi(q)$. The magnitude squared of the component of the state vector, i.e., $|\psi_n|^2$ or $|\psi(q)|^2$ gives the probability of the physical quantity to take on the n-th (or q-th) value when the quantity is measured. This is the physical interpretation (or significance) of the state vector. Therefore, each component of the state vector is called a *probability amplitude*. The familiar "wavefunction" in the Schrödinger formalism of wave mechanics is the probability amplitude where the physical quantities corresponding to the coordinate axes in linear space are the position coordinates and time, i.e.,

$$\psi(\vec{r}) = \psi(x, y, z, t) \ . \tag{2.3}$$

*A physical quantity, by definition, is anything that can be measured, even if by a gedanken experiment only.

In wave mechanics, the Schrödinger equation for a single particle tells us how the wavefunction evolves in time and space:

$$i\hbar\frac{\partial\psi(\vec{r})}{\partial t} = H_0\psi(\vec{r}) \ . \tag{2.4}$$

If we neglect spin, then

$$\begin{aligned} H_0 &= \frac{|\vec{p}|^2}{2m} + V(\vec{r}) \\ \vec{p} &= p_x\hat{\mathbf{x}} + p_y\hat{\mathbf{y}} + p_z\hat{\mathbf{z}} = -i\hbar\frac{\partial}{\partial x}\hat{\mathbf{x}} - i\hbar\frac{\partial}{\partial y}\hat{\mathbf{y}} - i\hbar\frac{\partial}{\partial z}\hat{\mathbf{z}}, \\ \vec{r} &= \left[x\hat{\mathbf{x}} + y\hat{\mathbf{y}} + z\hat{\mathbf{z}}\right], t \end{aligned} \tag{2.5}$$

where the quantities with "hats" are unit vectors along the coordinate axes.

Solution of Equation (2.4) yields the wavefunction $\psi(\vec{r})$. The quantity H_0 is the so-called Hamiltonian whose first term is the kinetic energy and the second term is the potential energy. The only restriction is that the potential energy term should be a real quantity so that the Hamiltonian remains a Hermitian operator, which guarantees that its eigenvalue (which is its expected value and therefore the expected value of the energy) remains a real quantity.

The question now is how to include "spin" in Equation (2.4)?

This was investigated by Wolfgang Pauli. He derived an equation to replace Equation (2.4) which bears his name and is known as the *Pauli Equation*. But before we discuss this equation, we need to understand an important concept, namely *Pauli spin matrices* since they appear in the Pauli Equation.

2.1 Pauli spin matrices

In quantum mechanics, any physical observable is associated with an operator (which would be a linear operator in the Schrödinger formalism, or a matrix in the Heisenberg formalism). The eigenvalues of the linear operator, or the eigenvalues of the matrix, are the expectation values of the physical quantity, i.e., the values we expect to find if we measure the physical quantity in an experiment[†]. Spin is a physical observable since the associated angular momentum can be measured, as was done unwittingly by Stern and Gerlach. Consequently, there must be a quantum mechanical operator associated with spin. Pauli derived the quantum mechanical operators for the spin components along three orthogonal axes – S_x, S_y and S_z. They are 2×2 complex

[†]Repeated measurements of a physical observable will produce a distribution of values whose average will be the expectation value.

matrices that came to be known as the Pauli spin matrices. Pauli's approach was based on the premise that: (1) the measurement of the spin angular momentum component along any coordinate axis for an electron should give the results $+\frac{\hbar}{2}$ or $-\frac{\hbar}{2}$, and (2) the operators for spin components along three mutually orthogonal axes should obey commutation rules similar to those obeyed by the operators associated with components of the orbital angular momentum. This would put spin angular momentum and orbital angular momentum on the same footing.

The operators (matrices) for the orbital angular momentum are known to satisfy the commutation relations

$$\begin{aligned} L_y L_z - L_z L_y &= i\hbar L_x, \\ L_z L_x - L_x L_z &= i\hbar L_y, \\ L_x L_y - L_y L_x &= i\hbar L_z, \end{aligned} \tag{2.6}$$

which merely reflect the fact that the orbital angular momenta along any two mutually orthogonal axes cannot be simultaneously measured with absolute precision unless the orbital angular momentum along a third axis, perpendicular to both the other two axes, vanishes.

Pauli adopted similar commutation relations for the spin angular momentum operators S_x, S_y and S_z:

$$\begin{aligned} S_y S_z - S_z S_y &= i\hbar S_x, \\ S_z S_x - S_x S_z &= i\hbar S_y, \\ S_x S_y - S_y S_x &= i\hbar S_z. \end{aligned} \tag{2.7}$$

Now, in the Stern-Gerlach experiment, assuming that the z-axis is the axis joining the south to north pole of the magnet, the observation of two traces on the photographic plate was interpreted as being caused by a spin angular momentum $\vec{S}$ whose z-component has two values $\pm\frac{\hbar}{2}$. Therefore, the matrix operator S_z must be (i) a 2×2 matrix (because such a matrix has *two* eigenvalues), and (ii) these eigenvalues must be $\pm\frac{\hbar}{2}$.

A 2×2 matrix that has eigenvalues of $\pm\frac{\hbar}{2}$ is the matrix

$$M_{2\times 2} = \frac{\hbar}{2}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{2.8}$$

This is not the only 2×2 matrix with eigenvalues $\pm\frac{\hbar}{2}$ – there could be many others – but this is the matrix that Pauli chose as a start for the operator S_z.

Next, he had to find appropriate matrices to serve as operators S_x and S_y. Pauli realized that since the choice of the z-axis as the axis joining the north and south poles of the magnet is completely arbitrary, the result of the Stern-Gerlach measurement should not be affected if he had chosen this axis to be the x- or y-axis, instead. This means that the expectation values of S_x and

S_y, i.e., their eigenvalues, should also be $\pm\frac{\hbar}{2}$. Moreover, all three matrices – S_x, S_y and S_z – must satisfy the commutation relations in Equation (2.7).

Pauli first defined three dimensionless matrices σ_x, σ_y and σ_z such that

$$\begin{aligned} S_x &= \frac{\hbar}{2}\sigma_x, \\ S_y &= \frac{\hbar}{2}\sigma_y, \\ S_z &= \frac{\hbar}{2}\sigma_z. \end{aligned} \tag{2.9}$$

Since S_x, S_y and S_z must have eigenvalues of $\pm\frac{\hbar}{2}$, it is obvious that the σ-matrices must have eigenvalues of ± 1. Furthermore, Equation (2.7) mandates that

$$\begin{aligned} \sigma_y\sigma_z - \sigma_z\sigma_y &= 2i\sigma_x, \\ \sigma_z\sigma_x - \sigma_x\sigma_z &= 2i\sigma_y, \\ \sigma_x\sigma_y - \sigma_y\sigma_x &= 2i\sigma_z. \end{aligned} \tag{2.10}$$

According to Equations (2.8) and (2.9),

$$\sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{2.11}$$

So now Pauli needed to pick two matrices σ_x and σ_y such that they have eigenvalues of ± 1 and obey Equation (2.10). Since these matrices will be operators for physical observables (spin components), they must be Hermitian as well. It is easy to verify that σ_z is Hermitian.

We can start our search for σ_x and σ_y with Hermitian matrices that have off-diagonal elements only, i.e.:

$$\sigma_x = \begin{pmatrix} 0 & a \\ a^* & 0 \end{pmatrix}, \tag{2.12}$$

and

$$\sigma_y = \begin{pmatrix} 0 & b \\ b^* & 0 \end{pmatrix}. \tag{2.13}$$

Since the eigenvalues of these matrices are ± 1, we must have $|a|^2 = |b|^2 = 1$, which leads to the possible choices for a and $b = \pm 1$ or $\pm i$.

Next, we must satisfy Equation (2.10) and that mandates

$$Im(ab^*) = 1, \tag{2.14}$$

where *Im* stands for imaginary part.

Therefore, if we select $a = +1$, then we must choose b = -i, and this yields

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \tag{2.15}$$

and

$$\sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}. \tag{2.16}$$

This is how Pauli came up with expressions for σ_x, σ_y and σ_z. These matrices are called *Pauli spin matrices* and serve as operators for the spin components according to Equation (2.9).

It is obvious that Pauli's choice was by no means unique. There are other legitimate choices (e.g., we could have chosen a = -i and $b = +1$), but Pauli's choice is now history and universally adopted.

From the expressions for the Pauli spin matrices, we notice that the square of each of the Pauli matrices is the 2×2 unit matrix $[I]$. Hence

$$|S|^2 = S_x^2 + S_y^2 + S_z^2 = \frac{3}{4}\hbar^2[I] = \overline{s}(\overline{s}+1)\hbar^2[I], \tag{2.17}$$

with $\overline{s} = 1/2$. This should be compared with the equivalent relation for the orbital angular momentum operator

$$|L|^2 = m(m+1)\hbar^2[I], \quad m = 1, 2, 3... \tag{2.18}$$

2.1.1 Eigenvectors of the Pauli matrices: spinors

The eigenvalues of the Pauli spin matrices are ±1. We now evaluate the corresponding eigenvectors that we denote as $|\pm>$.

Matrix σ_z: The eigenvectors of σ_z must satisfy

$$\sigma_z|\pm>_z = \pm 1|\pm>_z . \tag{2.19}$$

These eigenvectors (with unit norm) will be

$$|+>_z = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \tag{2.20}$$

and

$$|->_z = \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \tag{2.21}$$

It is easy to verify that these two eigenvectors are orthonormal, as they must be since they are eigenvectors of a Hermitian matrix corresponding to distinct (non-degenerate) eigenvalues.

Matrix σ_x: The eigenvectors of σ_x must satisfy

$$\sigma_x|\pm>_x = \pm 1|\pm>_x . \tag{2.22}$$

Starting with Equation (2.15), these eigenvectors are found to be

$$|+>_x = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \tag{2.23}$$

and

$$|- >_x = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix}. \tag{2.24}$$

Once again, the two eigenvectors are orthonormal. As can be easily checked, these eigenvectors can also be expressed as

$$|\pm >_x = \frac{1}{\sqrt{2}}[|+ >_z \pm |- >_z]. \tag{2.25}$$

Matrix σ_y: The eigenvectors of σ_y must satisfy

$$\sigma_y |\pm >_y = \pm 1 |\pm >_y . \tag{2.26}$$

Using Equation (2.16), these eigenvectors are found to be

$$|+ >_y = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ i \end{pmatrix}, \tag{2.27}$$

and

$$|- >_y = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -i \end{pmatrix}. \tag{2.28}$$

These eigenvectors are orthonormal and can be expressed as

$$|\pm >_y = \frac{1}{\sqrt{2}}[|+ >_z \pm i |- >_z]. \tag{2.29}$$

The eigenvectors of the Pauli spin matrices are examples of "spinors" which are 2×1 column vectors that represent the spin state of an electron. If we know the spinor associated with an electron in a given state, we can deduce the electron's spin orientation, i.e., find the quantities $< S_x >$, $< S_x >$ and $< S_x >$, where the angular brackets $< ... >$ denote expectation values. We will see this later.

2.2 The Pauli Equation and spinors

We can absorb the space and time dependent part of an electron's wavefunction in the spinor, so that the general form of a spinor will be

$$[\psi(\mathbf{x})] = \begin{bmatrix} \phi_1(\mathbf{x}) \\ \phi_2(\mathbf{x}) \end{bmatrix} , \tag{2.30}$$

where $\mathbf{x} \equiv (x, y, z, t)$, and ϕ_1 and ϕ_2 are the two components of the spinor wavefunction (assumed to be properly normalized).

With a 2-component wavefunction, the Schrödinger equation must be recast as

$$\left\{[H] + \frac{\hbar}{i}\frac{\partial}{\partial t}[I]\right\}[\psi(\mathbf{x})] = [0] \ , \tag{2.31}$$

where the Hamiltonian is a 2×2 matrix (since it may contain the 2×2 Pauli spin matrices), $[I]$ is the 2×2 identity matrix, and $[0]$ is the 2×1 null vector. Equation (2.31) is a set of two simultaneous differential equations for the two components of the spinor wavefunction – ϕ_1 and ϕ_2. Equation (2.31) is referred to as the *Pauli Equation*.

Solution of the Pauli Equation yields the two-component spinor wavefunction $[\psi(\mathbf{x})]$. Its practical use is in calculating the expected value of the spin angular momentum of an electron along any coordinate axis. The expected value along the n-th coordinate axis at location ($\vec{r}_0 = x_0, y_0, z_0$) at an instant of time t will be $[\psi(\vec{r}_0,t)]^\dagger[S_n][\psi(\vec{r}_0,t)]$, where $S_n = (\hbar/2)\sigma_n$ and the superscript † (dagger) represents the Hermitean conjugate. Using Equation (2.9), we get

$$\begin{aligned}
< S_x > (\vec{r}_0,t) &= (\hbar/2)[\psi(\vec{r}_0,t)]^\dagger\,[\sigma_x]\,[\psi(x_0,y_0,z_0,t)] \\
&= (\hbar/2)\,[\phi_1^*(\vec{r}_0,t)\ \phi_2^*(\vec{r}_0,t)] \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \phi_1(\vec{r}_0,t) \\ \phi_2(\vec{r}_0,t) \end{bmatrix} \\
&= \hbar Re\,[\phi_1^*(\vec{r}_0,t)\phi_2(\vec{r}_0,t)]\,, \\
< S_y > (\vec{r}_0,t) &= (\hbar/2)\,[\psi(\vec{r}_0,t)]^\dagger\,[\sigma_y]\,[\psi(\vec{r}_0,t)] \\
&= (\hbar/2)\,[\phi_1^*(\vec{r}_0,t)\ \phi_2^*(\vec{r}_0,t)] \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix} \begin{bmatrix} \phi_1(\vec{r}_0,t) \\ \phi_2(\vec{r}_0,t) \end{bmatrix} \\
&= \hbar Im\,[\phi_1^*(\vec{r}_0,t)\phi_2(\vec{r}_0,t)]\,, \\
< S_z > (\vec{r}_0,t) &= (\hbar/2)\,[\psi(\vec{r}_0,t)]^\dagger\,[\sigma_z]\,[\psi(\vec{r}_0,t)] \\
&= (\hbar/2)\,[\phi_1^*(\vec{r}_0,t)\ \phi_2^*(\vec{r}_0,t)] \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} \phi_1(\vec{r}_0,t) \\ \phi_2(\vec{r}_0,t) \end{bmatrix} \\
&= (\hbar/2)\{|\phi_1(\vec{r}_0,t)|^2 - |\phi_2(\vec{r}_0,t)|^2\}\ ,
\end{aligned} \tag{2.32}$$

where Re stands for the real part, Im stands for the imaginary part and the superscript * (asterisk) represents complex conjugate.

Therefore, if we can find the 2-component wavefunction in Equation (2.30) by solving the Pauli Equation (2.31), then we can find the three components of the expected value of the spin angular momentum at any location at any instant of time. This is why the Pauli Equation and the spinor concept are useful and important. In Chapter 7, we will also show how the Pauli Equation can be used to derive the energy dispersion relations (relation between the energy and the wavevector) of an electron in a solid in the presence of spin-dependent effects.

2.3 More on the Pauli Equation

Referring to Equation (2.31), we ask what terms will the 2×2 Hamiltonian $[H]$ contain. Normally, it will consist of three types of terms:

$$[H] = H_0[I] + [H_B] + [H_{SO}], \tag{2.33}$$

where H_0 is the spin-independent Hamiltonian, and $[H_B]$, $[H_{SO}]$ are 2×2 matrices that depend on spin and will therefore involve the Pauli spin matrices.

To understand where $[H_B]$ comes from, consider the fact that if we view "spin" as being associated with self rotation of an electron about its axis, then the self-rotation of the charged entity will give rise to a magnetic moment $\vec{\mu}_e$. This magnetic moment will interact with any externally applied magnetic field, if such a field is present. Let us say that the flux density associated with the external field is $\vec{B}$. Then the energy of interaction of $\vec{\mu}_e$ with $\vec{B}$ is

$$E_{int} = -\vec{\mu}_e \cdot \vec{B}. \tag{2.34}$$

Landé had shown that the ratio of the magnetic moment $\vec{\mu}_e$ (in units of the Bohr magneton μ_B) to the angular momentum of self rotation $\vec{S}$ (in units of $\hbar$) is the so-called gyromagnetic factor g [2]. Therefore, the operator associated with E_{int} is

$$[H_B] = -(g/2)\mu_B \vec{B} \cdot \vec{\sigma}, \tag{2.35}$$

since $\vec{S} = (\hbar/2)\vec{\sigma}$, where $\vec{\sigma} = \sigma_x \hat{x} + \sigma_y \hat{y} + \sigma_z \hat{z}$.

Obviously, the two eigenvalues of the matrix $[H_B]$ will not be the same, meaning that the eigenenergies associated with this Hamiltonian will *not* be degenerate. Therefore, this term will cause spin-splitting, or lift the degeneracy between the two spin states. This splitting is the *Zeeman splitting*. The Hamiltonian $[H_B]$ is called the *Zeeman Hamiltonian* or the *Zeeman interaction term*.

The Hamiltonian $[H_{SO}]$ is associated with spin-orbit interaction which also lifts the spin degeneracy. This interaction is discussed in Chapter 6.

Finally, the general form of the Pauli Equation is

$$\left\{ H_0[I] + [H_B] + [H_{SO}] + \frac{\hbar}{i}\frac{\partial}{\partial t}[I] \right\} [\psi(\mathbf{x})] = [0]. \tag{2.36}$$

Its solution yields the 2-component wavefunction $[\psi(\mathbf{x})]$, which then yields the spin components of an electron at any position and at any instant of time from Equation (2.32). The Pauli Equation of course has many other uses, as we will see later in Chapter 7.

2.4 Extending the Pauli Equation - the Dirac Equation

The Pauli Equation is completely non-relativistic and Pauli never found the avenue to reconcile it with relativity. That task was completed by Paul Andrew Maurice Dirac.

Both Schrödinger and two physicists O. Klein and W. Gordon had independently derived a relativistic equivalent of the Schrödinger equation. This is known as the *Klein-Gordon equation* [4]. A free particle not subjected to any force has a constant potential energy which can be taken to be zero (since potential is always undefined to the extent of an arbitrary constant). According to Einstein's special theory of relativity, such a particle obeys the relation

$$\overline{E}^2 = p^2c^2 + m_0^2c^4 \ , \tag{2.37}$$

where $\overline{E}$ is the total energy and p is the momentum.

According to De Broglie[‡],

$$\begin{aligned} \overline{E} &= h\nu, \\ p &= h/\lambda \ , \end{aligned} \tag{2.38}$$

where ν is the frequency and λ is the wavelength of the De Broglie wave associated with the particle ($\lambda = c\nu$). Therefore, the last equation can be re-written as

$$\nu^2 - \left(\frac{c}{\lambda}\right)^2 = \left(\frac{m_0c^2}{h}\right)^2 \ . \tag{2.39}$$

The above equation is known as the *Einstein-De Broglie Equation.*

In the operator version of quantum mechanics, the energy operator is $i\hbar(\partial/\partial t)$, and the momentum operator (describing momentum along the x_r axis) is $-i\hbar(\partial/\partial x_r)$. Therefore, the quantum mechanical representation of Equation (2.37) is

$$\left[\left(i\hbar\frac{\partial}{c\partial t}\right)^2 - \sum_{r=1}^{3}\left(-i\hbar\frac{\partial}{\partial x_r}\right)^2 - m_0^2c^2\right]\psi(x,y,z,t) = 0 \ . \tag{2.40}$$

The above equation has the solution of a plane wave

$$\begin{aligned} \psi(x,y,z,t) &= e^{i(\vec{k}\cdot\vec{r}-\omega t)}, \\ \omega &= 2\pi\nu, \\ \vec{k} &= (2\pi)/\vec{\lambda}, \\ \vec{r} &= x\hat{x} + y\hat{y} + z\hat{z} \ . \end{aligned} \tag{2.41}$$

[‡]The De Broglie relation has been reviewed in Chapter 15.

Substituting this solution into Equation (2.40) immediately yields Equation (2.39).

Equation (2.40) is valid for a free particle. Klein and Gordon extended it to a particle subjected to a (time-dependent) force field. Let the time dependent vector potential associated with the force field be $\vec{A} = (A_0, A_x, A_y, A_z)$, where we have treated space and time on the same footing as mandated by the theory of relativity. Klein and Gordon modified Equation (2.40) as

$$\left[\left(i\hbar\frac{\partial}{c\partial t} + eA_0\right)^2 - \sum_{r=1}^{3}\left(-i\hbar\frac{\partial}{\partial x_r} + eA_r\right)^2 - m_0^2c^2\right]\psi(x,y,z,t) = 0. \tag{2.42}$$

The above equation is the *Klein-Gordon Equation*. For a time, it was thought to be the fundamental equation of relativistic quantum mechanics.

Dirac however questioned the Klein-Gordon construct. This equation is a second order differential equation with respect to time and according to Dirac's transformation theory, all meaningful equations of quantum mechanics must be first order with respect to time (the Schrödinger and Pauli Equations are).

Dirac insisted on an equation that will be first order with respect to time. Now, in relativity, space and time are treated as equivalent and therefore the desired equation must also be first order with respect to space. Accordingly, the sought after equation needs to have a form

$$\left[\left(i\hbar\frac{\partial}{c\partial t} + eA_0\right) - \sum_{r=1}^{3}\alpha_r\left(-i\hbar\frac{\partial}{\partial x_r} + eA_r\right) - \alpha_0 m_0 c\right]\psi(x,y,z,t) = 0. \tag{2.43}$$

When Dirac postulated the above equation, he simultaneously came up with four new quantities α_0 and α_r (r = 1,2,3), and he also told us how to determine them. He realized that a free particle without the vector potentials A_0 and A_r has to satisfy the Einstein-De Broglie Equation (Equation(2.39)) and therefore must satisfy Equation (2.40). In other words, the wavefunction $\psi(x,y,z,t)$ must be a solution of Equation (2.40). The latter equation is second order in space and time, whereas Dirac's Equation (Equation(2.43)) for a free particle,

$$\left[\left(i\hbar\frac{\partial}{c\partial t}\right) - \sum_{r=1}^{3}\alpha_r\left(-i\hbar\frac{\partial}{\partial x_r}\right) - \alpha_0 m_0 c\right]\psi(x,y,z,t) = 0, \tag{2.44}$$

is first order.

In order to take the above Dirac's Equation for a free particle and make it second order to match Equation (2.40), we apply it to the operator

$$\left[\left(i\hbar\frac{\partial}{c\partial t}\right) + \sum_{r=1}^{3}\alpha_r\left(-i\hbar\frac{\partial}{\partial x_r}\right) + \alpha_0 m_0 c\right], \tag{2.45}$$

to yield

$$\left[\left(i\hbar\frac{\partial}{c\partial t}\right)+\sum_{r=1}^{3}\alpha_r\left(-i\hbar\frac{\partial}{\partial x_r}\right)+\alpha_0 m_0 c\right]$$
$$\cdot\left[\left(i\hbar\frac{\partial}{c\partial t}\right)-\sum_{r=1}^{3}\alpha_r\left(-i\hbar\frac{\partial}{\partial x_r}\right)-\alpha_0 m_0 c\right]\psi(x,y,z,t)=0. \quad (2.46)$$

Dirac insisted that the above equation be Equation (2.40). This can only happen if the quantities α_0 and α_r are not ordinary numbers, but *matrices*. In that case, Equation (2.46) will become

$$[\left(i\hbar\frac{\partial}{c\partial t}\right)^2-\sum_{r=1}^{3}\{\alpha_r\}^2\left(-i\hbar\frac{\partial}{\partial x_r}\right)^2-$$
$$\sum_{m<n}(\{\alpha_m\}\{\alpha_n\}-\{\alpha_n\}\{\alpha_m\})(i\hbar)^2\frac{\partial^2}{\partial x_m x_n}-\{\alpha_0\}^2 m_0^2 c^2]\psi(x,y,z,t)=0 \quad (2.47)$$

where the matrices $\{\alpha_0\}$ and $\{\alpha_r\}$ $(r = 1, 2, 3)$ have the properties

$$\{\alpha_m\}^2=[I] \qquad (m=0,1,2,3),$$
$$\{\alpha_m\}\{\alpha_n\}+\{\alpha_n\}\{\alpha_m\}=[0]\ (m\neq n; m,n=0,1,2,3). \quad (2.48)$$

The simplest matrices that possess the properties listed in Equation (2.48) are

$$\{\alpha_0\}=\begin{bmatrix}1&0&0&0\\0&1&0&0\\0&0&-1&0\\0&0&0&-1\end{bmatrix},$$
$$\{\alpha_1\}=\begin{bmatrix}0&0&0&1\\0&0&1&0\\0&1&0&0\\1&0&0&0\end{bmatrix},$$
$$\{\alpha_2\}=\begin{bmatrix}0&0&0&-i\\0&0&i&0\\0&-i&0&0\\i&0&0&0\end{bmatrix},$$
$$\{\alpha_3\}=\begin{bmatrix}0&0&1&0\\0&0&0&-1\\1&0&0&0\\0&-1&0&0\end{bmatrix}. \quad (2.49)$$

When these 4×4 matrices are introduced into Equation (2.47), the wavefunction $\psi(x, y, z, t)$ becomes a 4×1 column vector

$$\psi(x, y, z, t) = \begin{bmatrix} \psi_1 \\ \psi_2 \\ \psi_3 \\ \psi_4 \end{bmatrix} , \tag{2.50}$$

so that Equation (2.47) becomes a set of four coupled differential equations with respect to the four variables ψ_1, ψ_2, ψ_3 and ψ_4.

A little bit of inspection will quickly reveal that the Dirac matrices in Equation (2.49) can be written in terms of the Pauli spin matrices as follows

$$\begin{aligned} \{\alpha\}_0 &= \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & -\mathbf{I} \end{bmatrix}, \\ \{\alpha\}_1 &= \begin{bmatrix} \mathbf{0} & \sigma_1 \\ \sigma_1 & \mathbf{0} \end{bmatrix}, \\ \{\alpha\}_2 &= \begin{bmatrix} \mathbf{0} & \sigma_2 \\ \sigma_2 & \mathbf{0} \end{bmatrix}, \\ \{\alpha\}_3 &= \begin{bmatrix} \mathbf{0} & \sigma_3 \\ \sigma_3 & \mathbf{0} \end{bmatrix} , \end{aligned} \tag{2.51}$$

where $\mathbf{I}$ is a 2×2 identity matrix and $\mathbf{0}$ is a 2×2 null matrix.

Dirac applied his equation to the Hydrogen atom and showed that the correct atomic level spacings can be obtained. More importantly, he showed that the orbital angular momentum *alone* is not a conserved quantity, but when the quantity represented by the operator

$$\frac{1}{2} \begin{bmatrix} \mathbf{0} & \sigma \\ \sigma & \mathbf{0} \end{bmatrix} \tag{2.52}$$

is added to it, the total quantity is conserved. Viewed from the perspective of conservation of total angular momentum, this shows that an electron has spin angular momentum given by the operator in Equation (2.52). *This is the first convincing theoretical demonstration of the existence of spin* and hence Dirac is credited with establishing the concept of spin rigorously. In the process, he also demonstrated that spin angular momentum must be quantized to two distinct values since the matrix in Equation (2.52) has two distinct and discrete eigenvalues. Therefore, Dirac was able to explain spin quantization, which the self rotation model of the electron could never explain by itself.

Dirac also found that when an external force field is present, the procedure used to obtain Equation (2.47) from Equation (2.44) does not yield the Klein-Gordon equation (Equation(2.42)). The discrepancy can be explained by taking into account the interaction of the spin magnetic moment with the external field (recall the Zeeman interaction). Once this was demonstrated, the concept of spin was established on a firm footing.

2.5 The time independent Dirac Equation

From Equation (2.47), it is straightforward to show that the *time independent* Dirac Equation will be

$$\begin{bmatrix} A & 0 & C & D^* \\ 0 & A & D & -C \\ C & D^* & B & 0 \\ D & -C & 0 & B \end{bmatrix} \begin{pmatrix} \psi_1 \\ \psi_2 \\ \psi_3 \\ \psi_4 \end{pmatrix} = \overline{E} \begin{pmatrix} \psi_1 \\ \psi_2 \\ \psi_3 \\ \psi_4 \end{pmatrix}, \tag{2.53}$$

where the asterisk denotes complex conjugate and

$$\begin{aligned} A &= m_0c^2 + V \\ B &= -m_0c^2 + V \\ C &= c(p_z + eA_z) \\ D &= c[(p_x + eA_x) + i(p_y + eA_y)]. \end{aligned} \tag{2.54}$$

Equation (2.53) can be written more compactly as

$$\begin{bmatrix} (m_0c^2 + V)[I] & c\vec{\sigma} \cdot \left[\vec{p} + e\vec{A}\right] \\ c\vec{\sigma} \cdot \left[\vec{p} + e\vec{A}\right] & (-m_0c^2 + V)[I] \end{bmatrix} \begin{bmatrix} \{\psi\}(x,y,z) \\ \{\phi\}(x,y,z) \end{bmatrix} = \overline{E} \begin{bmatrix} \{\psi\}(x,y,z) \\ \{\phi\}(x,y,z) \end{bmatrix}, \tag{2.55}$$

where V is the scalar potential energy and

$$\{\psi\}(x,y,z) = \begin{pmatrix} \psi_1(x,y,z) \\ \psi_2(x,y,z) \end{pmatrix}, \tag{2.56}$$

and

$$\{\phi\}(x,y,z) = \begin{pmatrix} \psi_3(x,y,z) \\ \psi_4(x,y,z) \end{pmatrix}. \tag{2.57}$$

From Equation (2.55), we can show that

$$\begin{aligned} &\left\{ \left(m_0c^2 + V\right)[I] + \left[c\vec{\sigma} \cdot (\vec{p} + e\vec{A})\right] \frac{1}{\overline{E} + m_0c^2 - V} [I] \left[c\vec{\sigma} \cdot (\vec{p} + e\vec{A})\right] \right\} [\psi] = \overline{E}[\psi], \\ &\left\{ \left(-m_0c^2 + V\right)[I] + \left[c\vec{\sigma} \cdot (\vec{p} + e\vec{A})\right] \frac{1}{\overline{E} - m_0c^2 - V} [I] \left[c\vec{\sigma} \cdot (\vec{p} + e\vec{A})\right] \right\} [\phi] = \overline{E}[\phi]. \end{aligned} \tag{2.58}$$

2.5.1 Non-relativistic approximation to the Dirac Equation

Consider a non-relativistic electron moving at speeds much less than the speed of light in vacuum. For such a particle, Equation (2.37) yields $\overline{E} \approx m_0c^2$.

Using that result in the first of the two equations above yields

$$\left\{ \left(m_0c^2 + V\right)[I] + \frac{\left[\vec{\sigma}\cdot\left(\vec{p} + e\vec{A}\right)\right]^2}{2m_0} \right\}[\psi] = \overline{E}[\psi], \tag{2.59}$$

which reduces to

$$\begin{aligned}\left(\overline{E} - m_0c^2\right)[\psi] = E[\psi] &= \left(\frac{(\vec{p} + e\vec{A})^2}{2m_0}[I] + \mu_B\vec{B}\cdot\vec{\sigma} + V[I] \right)[\psi], \\ &= \{[H_0] + [H_B]\}[\psi],\end{aligned} \tag{2.60}$$

since $[\vec{\sigma}\cdot(\vec{p} + e\vec{A})]^2 = (\vec{p} + e\vec{A})^2[I] + 2m_0\mu_B\vec{B}\cdot\vec{\sigma}$ (see Problem 2.5). Here E is the total energy *minus* the rest energy m_0c^2.

The second of the two equations in Equation (2.58) will yield the same result as Equation (2.60) provided we make the transformation $m_0 \rightarrow -m_0$. This shows that the second equation applies to particles with negative mass, namely *anti-matter*. The existence of anti-matter is a foregone conclusion from Equation (2.37). Noting that De Broglie relation relates momentum p to wavevector k of a particle as $p = \hbar k$, Equation (2.37) gives *two* dispersion relations E versus k. They are shown in Fig. 2.1. One branch has a positive curvature and therefore positive mass. This corresponds to "matter". The other has a negative curvature and therefore negative mass. That corresponds to "anti-matter". The energy separation between the two curves is $2m_0c^2$ which is ~ 1 MeV for a free electron. These energy scales are seldom encountered in solid state physics, which is why anti-matter is usually of concern only in high energy physics. We will not have any occasion to worry about anti-matter anywhere in this text book.

2.5.2 Relationship between the non-relativistic approximation to the Dirac Equation and the Pauli Equation

The reader will immediately recognize Equation (2.60) as the Pauli Equation (Equation 2.33) without the spin-orbit interaction term. The latter term comes about from a different set of physics that has not been included here. However, what is amazing is that the Zeeman interaction term appears automatically, without having to introduce it separately. Therefore, Dirac was able to explain the Zeeman interaction directly from his equation!

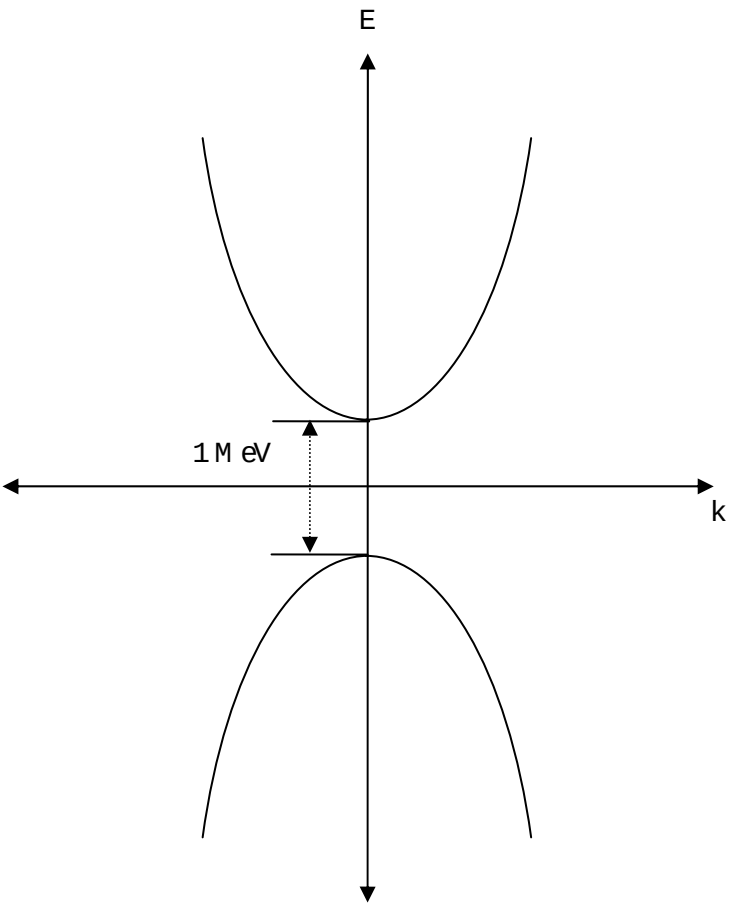


FIGURE 2.1
Dispersion relations for matter and anti-matter branches.

2.6 Problems

- **Problem 2.1**

  Show that the operators for S_x, S_y and S_z satisfy Equations (2.7) and (2.17).

  Solution

$$S_xS_y - S_yS_x = (\hbar^2/4)\left\{\begin{bmatrix}0&1\\1&0\end{bmatrix}\begin{bmatrix}0&-i\\i&0\end{bmatrix} - \begin{bmatrix}0&-i\\i&0\end{bmatrix}\begin{bmatrix}0&1\\1&0\end{bmatrix}\right\},$$
$$= \left(\frac{\hbar}{2}\right)^2\begin{bmatrix}2i&0\\0&-2i\end{bmatrix} = i\frac{\hbar}{2}\left[\frac{\hbar}{2}\sigma_z\right] = i\frac{\hbar}{2}S_z. \qquad (2.61)$$

  Furthermore,

$$|S|^2 = S_x^2 + S_y^2 + S_z^2,$$
$$= (\hbar^2/4)\left\{\begin{bmatrix}0&1\\1&0\end{bmatrix}^2 + \begin{bmatrix}0&-i\\i&0\end{bmatrix}^2 + \begin{bmatrix}1&0\\0&-1\end{bmatrix}^2\right\},$$
$$= (3/4)\hbar^2\begin{bmatrix}1&0\\0&1\end{bmatrix} = (3/4)\hbar^2[I]. \qquad (2.62)$$

- **Problem 2.2**

  Using Equation (2.32), show that if the 2-component spinor is the state $|+>_z$ given in Equation (2.20), then $< S_x > = < S_y > = 0$ and $< S_z > = \hbar/2$. Also show that if the 2-component spinor is the state $|->_z$ given in Equation (2.21), then $< S_x > = < S_y > = 0$ and $< S_z > =$ -$\hbar/2$. Hence the state $|+>_z$ is referred to as the *+z-polarized state* and the state $|->_z$ is referred to as the *-z-polarized state*. This means that an electron in these states has its spin polarized along the +z and -z axes, respectively.

  Solution

  For the state $|+>_z$, $\phi_1 = 1$ and $\phi_2 = 0$ in Equation (2.30).

  From Equation (2.32)

$$< S_x >= \hbar Re(\phi_1^*\phi_2) = 0,$$
$$< S_y >= \hbar Im(\phi_1^*\phi_2) = 0,$$
$$< S_z >= (\hbar/2)[|\phi_1|^2 - |\phi_2|^2] = \hbar/2. \qquad (2.63)$$

  For the state $|->_z$, $\phi_1 = 0$ and $\phi_2 = 1$, hence

$$< S_x >= \hbar Re(\phi_1^* \phi_2) = 0,$$
$$< S_y >= \hbar Im(\phi_1^* \phi_2) = 0,$$
$$< S_z >= (\hbar/2)[|\phi_1|^2 - |\phi_2|^2] = -\hbar/2. \tag{2.64}$$

Similarly, for the states $|\pm >_x$, $< S_x >= \pm\frac{\hbar}{2}$ and $< S_y >=< S_z >= 0$, so that these states are $\pm$x-polarized states. One can show that the states $|\pm >_y$ are $\pm$y-polarized states.

- **Problem 2.3**

  Show that if any 2×2 Hermitian matrix $[H]$ is used to perform a unitary transformation of the Pauli matrices defined above, i.e., if matrices $\sigma_x{}'$, $\sigma_y{}'$, $\sigma_z{}'$ are defined such that

  $$\sigma_n{}' = S\sigma_n S^{-1}, \tag{2.65}$$

  where $S = e^{iH}$, for $n = x$, y, or z, the commutation rules in Equation (2.7) are satisfied by the matrices $\sigma_n{}'$.

  Furthermore, the square of each of the matrices $\sigma_n{}'$ is equal to the 2×2 identity matrix.

  Solution

  The solution of this problem is left to the reader. Use the fact that the matrix S is

  $$[S] = [M]^{-1} \begin{bmatrix} e^{i\lambda_1} & 0 \\ 0 & e^{i\lambda_2} \end{bmatrix} [M], \tag{2.66}$$

  where $\lambda_{1,2}$ are the eigenvalues of the matrix $[H]$, and $[M]$ is a 2×2 matrix whose two columns are the eigenfunctions of $[H]$ corresponding to eigenvalues λ_1 and λ_2.

- **Problem 2.4**

  Show the following properties of the Pauli spin matrices by actual computation.

  1. det $(\sigma_j) = -1$; for j = x, y, or z.
  2. Tr $(\sigma_j) = 0$.
  3. $\sigma_x^2 = \sigma_y^2 = \sigma_z^2 = I$.
  4. $\sigma_x\sigma_y\sigma_z = iI$.
  5. $\sigma_x\sigma_y = -\sigma_y\sigma_x = i\sigma_z$.
  6. $\sigma_y\sigma_z = -\sigma_z\sigma_y = i\sigma_x$.
  7. $\sigma_z\sigma_x = -\sigma_x\sigma_z = i\sigma_y$.
  8. $\sigma_p\sigma_q + \sigma_q\sigma_p = 0, \quad (p \neq q; p, q = x, y, z)$.

- **Problem 2.5**

  Show that $\left[\vec{\sigma}\cdot(\vec{p}+e\vec{A})\right]^2 = (\vec{p}+e\vec{A})^2[I] + 2m_0\mu_B\vec{B}\cdot\vec{\sigma}$.

  Solution

$$\begin{aligned}\left[\vec{\sigma}\cdot(\vec{p}+e\vec{A})\right]^2 &= [\sigma_x(p_x+eA_x)+\sigma_y(p_y+eA_y)+\sigma_z(p_z+eA_z)]\\ &\quad [\sigma_x(p_x+eA_x)+\sigma_y(p_y+eA_y)+\sigma_z(p_z+eA_z)]\,,\\ &= \sigma_x^2(p_x+eA_x)^2+\sigma_y^2(p_y+eA_y)^2+\sigma_z^2(p_z+eA_z)^2\\ &\quad +\sigma_x\sigma_y(p_x+eA_x)(p_y+eA_y)+\sigma_y\sigma_x(p_y+eA_y)(p_x+eA_x)\\ &\quad \sigma_y\sigma_z(p_y+eA_y)(p_z+eA_z)+\sigma_z\sigma_y(p_z+eA_z)(p_y+eA_y)\\ &\quad \sigma_z\sigma_x(p_z+eA_z)(p_x+eA_x)+\sigma_x\sigma_z(p_x+eA_x)(p_z+eA_z).\end{aligned} \tag{2.67}$$

  Now the results of Problem 2.4 lead to

$$\begin{aligned}\left[\vec{\sigma}\cdot(\vec{p}+e\vec{A})\right]^2 &= (p_x+eA_x)^2+(p_y+eA_y)^2+(p_z+eA_z)^2\\ &\quad +i\sigma_z\left(-ie\hbar\frac{\partial A_y}{\partial x}+ie\hbar\frac{\partial A_x}{\partial y}\right)\\ &\quad +i\sigma_y\left(-ie\hbar\frac{\partial A_x}{\partial z}+ie\hbar\frac{\partial A_z}{\partial x}\right)\\ &\quad +i\sigma_x\left(-ie\hbar\frac{\partial A_z}{\partial y}+ie\hbar\frac{\partial A_y}{\partial z}\right)\\ &= (p_x+eA_x)^2+(p_y+eA_y)^2+(p_z+eA_z)^2\\ &\quad +e\hbar\sigma_z\left(\frac{\partial A_y}{\partial x}-\frac{\partial A_x}{\partial y}\right)\\ &\quad +e\hbar\sigma_y\left(\frac{\partial A_x}{\partial z}-\frac{\partial A_z}{\partial x}\right)\\ &\quad +e\hbar\sigma_x\left(\frac{\partial A_z}{\partial y}-\frac{\partial A_y}{\partial z}\right)\\ &= \left[\vec{p}+e\vec{A}\right]^2+e\hbar\left(\vec{\nabla}\times\vec{A}\right)\cdot\vec{\sigma},\end{aligned} \tag{2.68}$$

  where we have used the fact that $\vec{p}_n$ = -$i\hbar(\partial/\partial\vec{x}_n)$.

  Since $\mu_B = e\hbar/2m_0$ and $\vec{B} = \vec{\nabla}\times\vec{A}$, this proves the equality stated in the problem.

2.7 Appendix

2.7.1 Working with spin operators

Applying the operators σ_x, σ_y and σ_z to the 2-component wavefunction in Equation (2.30) yields

$$\begin{aligned}
\sigma_x\psi(\mathbf{x}) &= \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \phi_1(\mathbf{x}) \\ \phi_2(\mathbf{x}) \end{bmatrix}, \\
\sigma_y\psi(\mathbf{x}) &= \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix} \begin{bmatrix} \phi_1(\mathbf{x}) \\ \phi_2(\mathbf{x}) \end{bmatrix}, \\
\sigma_z\psi(\mathbf{x}) &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \phi_1(\mathbf{x}) \\ \phi_2(\mathbf{x}) \end{bmatrix},
\end{aligned} \tag{2.69}$$

and therefore

$$\begin{aligned}
S_x\psi &= (\hbar/2)\sigma_x\psi = \frac{\hbar}{2} \begin{bmatrix} \phi_2(\mathbf{x}) \\ \phi_1(\mathbf{x}) \end{bmatrix}, \\
S_y\psi &= (\hbar/2)\sigma_x\psi = \frac{\hbar}{2} \begin{bmatrix} -i\phi_2(\mathbf{x}) \\ i\phi_1(\mathbf{x}) \end{bmatrix}, \\
S_z\psi &= (\hbar/2)\sigma_x\psi = \frac{\hbar}{2} \begin{bmatrix} \phi_1(\mathbf{x}) \\ -\phi_2(\mathbf{x}) \end{bmatrix}.
\end{aligned} \tag{2.70}$$

Thus, operating with S_x interchanges the two components of the spinor, operating with S_y interchanges the two components of the spinor while causing a phase shift of – 90° to the second component and a phase shift of 90° to the first component, and operating with S_z introduces a phase shift of 180° to the second component. These results can be useful as we will show in Chapter 9.

2.7.2 Two useful theorems

A trivial decomposition of any 2×2 matrix M is obviously

$$M = m_{11} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + m_{12} \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} + m_{21} \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} + m_{22} \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \tag{2.71}$$

since the 4 matrices on the right hand side form a complete basis for all 2×2 matrices. Now that we have introduced the Pauli spin matrices, a more subtle decomposition of any 2×2 complex matrix can be found, as discussed next.

Theorem I: Any 2×2 matrix M

$$M = \begin{pmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{pmatrix} \tag{2.72}$$

can be decomposed as follows

$$M = \frac{m_{11}+m_{22}}{2}I + \frac{m_{11}-m_{22}}{2}\sigma_z + \frac{m_{12}+m_{21}}{2}\sigma_x + \frac{m_{12}-m_{21}}{2}\sigma_y. \quad (2.73)$$

The proof of which is left as an exercise. In other words, the 4 matrices $(I, \sigma_x, \sigma_y, \sigma_z)$ form a complete set of bases in the space of 2×2 complex matrices.

The last equation can be written in the more condensed form

$$M = a_o I + \vec{a} \cdot \vec{\sigma}, \quad (2.74)$$

where

$$a_o = \frac{1}{2}Tr(M), \quad (2.75)$$

and

$$\vec{a} = \frac{1}{2}Tr(M\vec{\sigma}), \quad (2.76)$$

where $\vec{\sigma} = (\sigma_x, \sigma_y, \sigma_z)$, and Tr stands for the trace of the matrix.

A comparison of Equations (2.73) and (2.74) shows that M is Hermitian if a_0 and the three components of the vector $\vec{a}$ are real.

Exercise: Calculate the values of $(a_0, \vec{a})$ for the matrix M given by

$$M = \begin{pmatrix} 2 & \frac{i\sqrt{3}}{3} \\ \frac{-i\sqrt{3}}{3} & 4 \end{pmatrix}. \quad (2.77)$$

Next, we prove an identity which will be used in the next chapter to interpret geometrically the Pauli matrices after the introduction of the Bloch sphere concept.

Theorem II: If θ is real and if the matrix A is such that $A^2 = I$, the following identity holds:

$$e^{i\theta A} = cos\theta I + isin\theta A. \quad (2.78)$$

This is the generalization to operators of the well-known Euler relation for complex numbers, i.e., $e^{iz} = cosz + isinz$.

From the Taylor series expansion

$$e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}, \quad (2.79)$$

and the definition of the function of an operator, we get

$$e^{i\theta A} = I + (i\theta)A + \frac{(i\theta)^2 A^2}{2!} + \frac{(i\theta)^3 A^3}{3!} + \frac{(i\theta)^4 A^4}{4!} + ... \quad (2.80)$$

or

$$e^{i\theta A} = (1 - \frac{\theta^2}{2!} + \frac{\theta^4}{4!} - ... + (-1)^k \frac{\theta^{2k}}{(2k)!})I$$
$$+i(\theta - \frac{\theta^3}{3!} + \frac{\theta^5}{5!} - ... + (-1)^k \frac{i\theta^{2k+1}}{(2k+1)!})A, \quad (2.81)$$

which is indeed Equation (2.78) if we use the the Taylor expansions

$$sinx = \sum_{k=0}^{\infty} (-1)^k x^{2k+1}/(2k+1)!, \quad (2.82)$$

and

$$cosx = \sum_{k=0}^{\infty} (-1)^k x^{2k}/(2k)!. \quad (2.83)$$

2.7.3 Applications of the *Postulates of Quantum Mechanics* to a few spin problems

The Postulates of Quantum Mechanics are briefly reviewed in Chapter 15.

Example 1: If we measure the z-component of an electron's spin, apply the postulate of quantum projective measurement (Postulate 3) discussed in Chapter 15 to calculate the probability of the measurement to give the result $\frac{\hbar}{2}$ or $-\frac{\hbar}{2}$, respectively, if prior to the measurement, the state of the system $|\psi>$ is either

(1) $|0>$,
(2) $|1>$, or
(3) $\frac{1}{\sqrt{2}}(|0> + |1>)$,

where $|0>$ is the +z-polarized state $|+>_z$ and $|1>$ is the -z-polarized state $|->_z$.

Solution

(1) The operator $S_z = \frac{\hbar}{2}\sigma_z$ has the following spectral decomposition

$$S_z = \frac{\hbar}{2}|0><0| + (-\frac{\hbar}{2})|1><1| = \sum_m mP_m, \quad (2.84)$$

where $P_m = |m><m|$, with $|m>$ being the eigenvectors of S_z.

Hence, if $|\psi>=|0>$, the probability to measure $+\frac{\hbar}{2}$ is equal to

$$p(+\frac{\hbar}{2}) =<0|(|0><0|)|0>= 1, \quad (2.85)$$

and the probability to measure $-\frac{\hbar}{2}$ is equal to

$$p(-\frac{\hbar}{2}) =< 0|(|1><1|)|0>= 0 \tag{2.86}$$

and the sum of the probabilities is indeed equal to unity.

(2) If $|\psi>=|1>$, we get

$$p(\frac{\hbar}{2}) =< 1|(|0><0|)|1>= 0, \tag{2.87}$$

and

$$p(-\frac{\hbar}{2}) =< 1|(|1><1|)|1>= 1. \tag{2.88}$$

(3) Finally, if $|\psi>=\frac{1}{\sqrt{2}}(|0>+|1>)$, we find

$$p(+\frac{\hbar}{2}) = \frac{1}{\sqrt{2}}\{|0>+<1|\}(|0><0|)\frac{1}{\sqrt{2}}\{|0>+<1|\} = \frac{1}{2}\{1+0\}\cdot\{0+1\} = \frac{1}{2}, \tag{2.89}$$

and similarly, $p(-\frac{\hbar}{2}) = \frac{1}{2}$.

According to *Postulate 3* discussed in Chapter 15, right after the measurement, the spinor collapses into the state

$$|\psi>\rightarrow|\psi^{new}>=\frac{(|0><0|)}{\sqrt{p(+\frac{\hbar}{2})}}|\psi>=(\frac{|0><0|}{\sqrt{\frac{1}{2}}})(\frac{(|0>+|1>)}{\sqrt{2}}) = |0>. \tag{2.90}$$

Exercise: Repeat the previous exercise if the component S_x is measured instead.

Example 2: Suppose an electron is prepared in the spinor $|0>$ eigenstate of S_z with eigenvalue $+\frac{\hbar}{2}$ and repeated measurements are made of the x-component of its intrinsic angular momentum, calculate the average value $< S_x >$ and the standard deviation $\Delta(S_x)$ of these measurements.

Solution

The spectral decomposition of the operator S_x is given by

$$S_x = \frac{\hbar}{2}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = (+\frac{\hbar}{2})|+>_{x,x}<+| + (-\frac{\hbar}{2})|->_{x,x}<-|, \tag{2.91}$$

where $|+>_x$ is the +x-polarized state and $|->_x$ is the -x-polarized state.

Hence,

$$< S_x >= \frac{\hbar}{2}\left[< 0|S_x|0 >= (1\ 0)\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 1 \\ 0 \end{pmatrix}\right] = 0. \tag{2.92}$$

Furthermore,

$$\begin{aligned} \Delta(S_x) &= \sqrt{< {S_x}^2 > - < S_x >^2} = \sqrt{< {S_x}^2 >}, \\ &= \frac{\hbar}{2}\left[< 0|\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}|0 >\right]^{1/2}, \\ &= \frac{\hbar}{2}\left[(1\ 0)\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 1 \\ 0 \end{pmatrix}\right]^{1/2} = \frac{\hbar}{2}. \end{aligned} \tag{2.93}$$

Example 3: Suppose an electron is characterized by the spinor

$$|\psi >= \frac{1}{\sqrt{10}}\begin{pmatrix} 3 \\ -1 \end{pmatrix}, \tag{2.94}$$

which is properly normalized, as easily checked. If we measure the y-component of the spin, what is the probability to find that its value is $\frac{\hbar}{2}$?

Solution

The eigenvector of σ_y are $\frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ i \end{pmatrix}$ with eigenvalue +1 and $\frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -i \end{pmatrix}$ with eigenvalue -1

Hence, if we write the spinor $|\psi >$ as a linear combination of $|+ >_y$ and $|- >_y$,

$$|\psi >= \frac{1}{\sqrt{10}}\begin{pmatrix} 3 \\ -1 \end{pmatrix} = \alpha\frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ i \end{pmatrix} + \beta\frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -i \end{pmatrix}. \tag{2.95}$$

The probability to find $+\frac{\hbar}{2}$ when measuring S_y is given by $|\alpha|^2$ or

$$p(+\frac{\hbar}{2}) = \left|_y < +|\psi > \right|^2 = |\frac{1}{\sqrt{2}}(1,\ -i)\frac{1}{\sqrt{10}}\begin{pmatrix} 3 \\ -1 \end{pmatrix}|^2 = \frac{1}{20}|(3+i)\Big|^2 = \frac{1}{2}. \tag{2.96}$$

Example 4: Suppose an electron is characterized by the spinor

$$|\psi >= \frac{4}{5}|0 > + \frac{3}{5}|1 > . \tag{2.97}$$

(1) What is the probability that a measurement of z-component of the spin will be $+\frac{\hbar}{2}$? and $-\frac{\hbar}{2}$?

Solution

$$p(+\frac{\hbar}{2}) = \left|< 0|(\frac{4}{5}|0 > + \frac{3}{5}|1 >)\right|^2 = \frac{16}{25}, \tag{2.98}$$

and

$$p(\frac{-\hbar}{2}) = \left| < 1|(\frac{4}{5}|0> + \frac{3}{5}|1>) \right|^2 = \frac{9}{25}. \tag{2.99}$$

(2) What is the expectation value of $S_z = \frac{\hbar}{2}\sigma_z$?

Solution

$$\begin{aligned} < \psi|S_z|\psi > &= \left(\frac{4}{5} < 0| + \frac{3}{5} < 1|\right) \frac{\hbar}{2}\sigma_z \left(\frac{4}{5}|0> + \frac{3}{5}|1>\right), \\ &= \frac{16}{25} < 0|\sigma_z|0> \frac{\hbar}{2} + \left(\frac{3}{5}\right)^2 \frac{\hbar}{2} < 1|\sigma_z|1 >= \frac{16}{25}\left(\frac{\hbar}{2}\right) - \frac{9}{25}\left(\frac{\hbar}{2}\right), \\ &= \frac{7}{25}\left(\frac{\hbar}{2}\right). \end{aligned} \tag{2.100}$$

(3) What is the standard deviation of S_z if measurements are made on many electrons prepared in the state $|\psi>$ above?

Solution
By definition

$$\triangle(S_z) = \sqrt{< \psi|{S_z}^2|\psi > - < S_z >^2}. \tag{2.101}$$

Since

$$< {S_z}^2 >=< \psi|{S_z}^2|\psi >=< \psi|\frac{\hbar^2}{4}{\sigma_z}^2|\psi >= \frac{\hbar^2}{4}, \tag{2.102}$$

and from the previous problem

$$\langle S_z \rangle = \frac{7}{25}\left(\frac{\hbar}{2}\right), \tag{2.103}$$

we get

$$\triangle(S_z) = \sqrt{\frac{\hbar^2}{4}(1 - \frac{49}{625})} = \frac{12}{25}\hbar. \tag{2.104}$$

Exercise: Proceeding as above, calculate the standard deviations $\triangle(S_x)$ and $\triangle(S_y)$.

Answer:

$$\Delta(S_x) = \frac{7}{25}\frac{\hbar}{2}, \tag{2.105}$$

and

$$\Delta(S_y) = \frac{\hbar}{2}. \tag{2.106}$$

2.7.4 The Heisenberg Principle for spin components

In Chapter 15, we prove the following general Heisenberg inequality

$$\triangle(C)\triangle(D) \geq \frac{|<\psi|[C,D]|\psi>|}{2}, \tag{2.107}$$

where $\triangle(C)$ and $\triangle(D)$ are the standard deviations associated with measurements of the observables C and D on a electron prepared many times in the same state $|\psi>$.

If we take as observables the two components of the intrinsic angular momentum of the electron σ_x and σ_y, the inequality above becomes

$$\triangle(S_x)\triangle(S_y) \geq \frac{\hbar}{2}|<\psi|S_z|\psi>| = \frac{\hbar^2}{4}|<\psi|\sigma_z|\psi>|. \tag{2.108}$$

Exercise: Prove that, in the last inequality, the equality signs holds when the state of the spinor is given by

$$|\psi>= \frac{3}{\sqrt{10}}|0> + \frac{-1}{\sqrt{10}}|1>, \tag{2.109}$$

considered in Example 3.

Solution

For an electron prepared in the state $|\psi>$ above, we use the results of Example 3 and find right-hand-side of Equation (2.108) to be equal to

$$\frac{\hbar^2}{4}|<\psi|\sigma_z|\psi>| = \frac{\hbar^2}{5}. \tag{2.110}$$

The left hand side of the inequality is

$$\triangle(S_x)\triangle(S_y) = (\frac{2}{5}\hbar)(\frac{\hbar}{2}) = \frac{\hbar^2}{5}. \tag{2.111}$$

Therefore, the inequality (2.108) reduces to an equality in this case.

Exercise: The following inequalities also hold

$$\triangle(S_y)\triangle(S_z) \geq \frac{\hbar}{2}|<\psi|S_x|\psi>| = \frac{\hbar^2}{4}|<\psi|\sigma_x|\psi>|, \tag{2.112}$$

and

$$\triangle(S_x)\triangle(S_z) \geq \frac{\hbar}{2}|<\psi|S_y|\psi>| = \frac{\hbar^2}{4}|<\psi|\sigma_y|\psi>|. \tag{2.113}$$

Prove that these inequalities are indeed satisfied for a spinor in the state (2.109) given above. The solution of this exercise is left for the reader.

2.8 References

[1] C. Ekart, "Operator Calculus and the Solution of the Equations of Dynamics", Phys. Rev., **28**, 711 (1926).

[2] R.P. Feynman, R.B. Leighton and M. Sands, *The Feynman Lectures on Physics*, Volume II, chapter 34, Addison-Wesley, Reading, Mass., (1965).

[3] P. A. M. Dirac, *The Principles of Quantum Mechanics*, Oxford University Press (1958).

[4] J. J. Sakurai, *Advanced Quantum Mechanics*, Addison-Wesley, Reading, Mass. (1967).

3

The Bloch Sphere

The Bloch sphere concept is a useful tool to represent the actions of various quantum-mechanical operators on a spinor. Most importantly, it provides a link between the rather abstract concept of a spinor and the more intuitive way (although not rigorous) of thinking of the spin of an electron as being associated with a magnetic moment along the axis of self rotation. The Bloch sphere is particularly useful when describing the action of a spatially uniform external (including time-dependent) magnetic field on the spin [1]. This will be illustrated in the next chapter through the derivation of the Rabi formula. Bloch sphere concepts are also frequently invoked in discussions of spin-based quantum computing, so that a sound understanding of this notion is imperative to comprehend the literature in that field.

3.1 The spinor and the "qubit"

In Chapter 14, where we discuss spintronic implementations of quantum computers and logic gates, we will encounter the notion of "qubit". The term "qubit" is a short form for "quantum bit". Quantum computers do not process classical binary bits 0 and 1, but instead process *quantum bits* which are coherent superpositions of both 0 and 1. A quantum computer derives its immense power from this superposition.

It was realized quite some time back that the spin polarization of an electron is the ideal physical implementation of a qubit since spin polarization can survive in a coherent superposition of two orthogonal states for a relatively long time. These two orthogonal states (i.e., states with anti-parallel spin polarizations) will represent the classical bits 0 and 1. We elucidate this below.

The 2-component wavefunction representing an arbitrary spin state can be written as

$$[\psi(\mathbf{x})] = \begin{bmatrix} \phi_1 \\ \phi_2 \end{bmatrix} = \phi_1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \phi_2 \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \phi_1 |+>_z + \phi_2 |->_z, \tag{3.1}$$

where

$$|\phi_1|^2 + |\phi_2|^2 = 1 \tag{3.2}$$

if $[\psi(\mathbf{x})]$ is properly normalized.

Therefore, any spin state can be expressed as a coherent superposition of the +z and -z-polarized states. The two z-polarized states, being mutually orthogonal, can represent the classical bits 0 and 1. The quantities ϕ_1 and ϕ_2 are complex quantities with both a magnitude and a phase. To maintain a *coherent* superposition state, we have to maintain the phase relationship between ϕ_1 and ϕ_2 until the qubit is "read" and collapses to either the +z- or the -z-polarized state*.

With "spin", the coherent superposition can be maintained for a much longer duration than with "charge". Spin couples relatively weakly to its surroundings. As a result, the coupling to the environment does not destroy the phase coherence of spin (phase relationship between ϕ_1 and ϕ_2) as rapidly as it does in the case of charge. This is a major advantage of spin over charge and it has made "spin" the preferred vehicle for implementing qubits in solid state quantum computers.

Since the +z-polarized state was assumed to represent the classical bit 0, we will denote the corresponding spinor as $|0>$ using the Dirac ket notation. It is given by

$$|0>=\begin{bmatrix}1\\0\end{bmatrix}. \tag{3.3}$$

Similarly, the spinor for the -z-polarized state will be denoted as the ket

$$|1>=\begin{bmatrix}0\\1\end{bmatrix}. \tag{3.4}$$

We can re-write Equation (3.1) as

$$|\chi>=\begin{bmatrix}\alpha\\\beta\end{bmatrix}=\alpha|0>+\beta|1>, \tag{3.5}$$

where $|\chi>$ is an arbitrary spinor and α, β are complex numbers whose square magnitudes denote the probability that if a measurement of the spin component along the z-axis is performed on the electron, it will be found to be in state $|0>$ with probability $|\alpha|^2$ and in state $|1>$ with probability $|\beta|^2$. Therefore, we must have

$$|\alpha|^2+|\beta|^2=1. \tag{3.6}$$

The above two equations immediately suggest that the spinor $\chi>$ can be viewed as a qubit since it has the exact mathematical form of a qubit, namely a coherent superposition of bits 0 and 1.

The quantities α, β are *continuous* variables whose magnitudes can take any value between 0 and 1 and whose phases can also take arbitrary values between 0 and 2π. On the other hand, the states $|0>$ and $|1>$ correspond

*The qubit will collapse to the +z-polarized state with probability $|\phi_1|^2$ and to -z-polarized state with probability $|\phi_2|^2$.

to digital (*discrete*) binary bits. Thus, quantum computation is neither quite analog computing, nor digital, but something in between.

Since complex numbers have a nice powerful geometrical (or Gauss) representation in the complex plane, one might ask if there is such a useful representation for the spinors or qubits in $\mathcal{C}^2$. The next sections will show that the *Bloch sphere* is precisely this representation.

3.2 The Bloch sphere concept

3.2.1 Preliminaries

In Chapter 2, we introduced the concept that an electron's spin is described by the operator

$$\vec{S}_{op} = \frac{\hbar}{2}\vec{\sigma}, \tag{3.7}$$

whose three components (S_x, S_y, S_z) satisfy commutation rules similar to those for the orbital angular momentum $\vec{L}$ (see Equation (2.7)). A measurement of the spin component along an arbitrary direction characterized by a unit vector $\hat{n}$ will yield results given by the eigenvalues of the operator

$$\vec{S}.\hat{n}, \tag{3.8}$$

and these eigenvalues are $\pm\frac{\hbar}{2}$, irrespective of the direction of the unit vector $\hat{n}$.

We prove this last statement starting with the following equality for the Pauli spin matrices

$$(\vec{\sigma}\cdot\vec{a})(\vec{\sigma}\cdot\vec{b}) = i\vec{\sigma}\cdot(\vec{a}\times\vec{b}) + \vec{a}\cdot\vec{b}I, \tag{3.9}$$

where I is the 2×2 identity matrix and $\vec{a}$ and $\vec{b}$ are any arbitrary three-dimensional vectors in real space $\Re^3$. The above identity is very easy to verify and is left as an exercise for the reader.

If the vectors $\vec{a}$ and $\vec{b}$ are equal to a unit vector $\hat{n}$, then the equality above reduces to

$$(\vec{\sigma}\cdot\hat{n})^2 = I, \tag{3.10}$$

i.e., the square of any component of $\vec{\sigma}$ is equal to the unit 2×2 matrix. Hence, the eigenvalues of $\vec{\sigma}.\hat{n}$ are ±1, and therefore (recall Equation (3.7)) the eigenvalues of the operator $\vec{S}.\hat{n}$ must be $\pm\hbar/2$, which proves the result we were after. In other words, the measurement of the spin angular momentum along any arbitrary axis always yields the values $\pm\hbar/2$.

3.2.1.1 Eigenvectors of $(\vec{\sigma}\cdot\hat{n})$

Let us derive the explicit analytical expressions for the eigenvectors of $\vec{\sigma}\cdot\hat{n}$ corresponding to the eigenvalues +1 and -1.

To achieve that goal, we consider the following operators

$$\frac{1}{2}(1\pm\vec{\sigma}\cdot\hat{n}), \tag{3.11}$$

acting on an arbitrary spinor or qubit $|\chi>$. If we operate on that with the operator $(\vec{\sigma}\cdot\hat{n})$, we get

$$(\vec{\sigma}\cdot\hat{n})\left[\frac{1}{2}(1\pm\vec{\sigma}\cdot\hat{n})|\chi>\right]=\frac{1}{2}\vec{\sigma}\cdot\hat{n}|\chi>\pm\frac{1}{2}\left(\vec{\sigma}\cdot\hat{n}\right)^2|\chi>=\pm\left[\frac{1}{2}(1\pm\vec{\sigma}\cdot\hat{n})|\chi>\right]. \tag{3.12}$$

This means that, for any $|\chi>$, $\frac{1}{2}(1\pm\vec{\sigma}\cdot\hat{n})|\chi>$ are eigenvectors of $\vec{\sigma}.\hat{n}$ with eigenvalues ±1. Making use of the identity

$$\frac{1}{2}(1\pm\vec{\sigma}.\hat{n})=\frac{1}{2}[1\pm\sigma_z n_z\pm\frac{1}{2}(\sigma_x+i\sigma_y)(n_x-in_y)\pm\frac{1}{2}(\sigma_x-i\sigma_y)(n_x+in_y)], \tag{3.13}$$

where n_x, n_y and n_z are the x-, y- and z-components of the vector $\hat{n}$, and using spherical coordinates with polar angle θ and azimuthal angle ϕ, so that

$$(n_x,n_y,n_z)=(sin\theta cos\phi, sin\theta sin\phi, cos\theta), \tag{3.14}$$

we get

$$n_x\pm in_y=sin\theta e^{\pm i\phi}, \tag{3.15}$$

which leads to

$$\frac{1}{2}(1\pm\vec{\sigma}\cdot\hat{n})=\frac{1}{2}\left[1\pm cos\theta\sigma_z\pm\frac{1}{2}(sin\theta e^{-i\phi}\sigma_+\pm sin\theta e^{i\phi}\sigma_-)\right], \tag{3.16}$$

where the operators σ_+ and σ_- are given by $\sigma_+=\sigma_x+i\sigma_y$ and $\sigma_-=\sigma_x-i\sigma_y$, respectively.

Acting with these operators on the ket $|0>$, we get

$$\frac{1}{2}(1+\vec{\sigma}\cdot\hat{n})|0>=cos\frac{\theta}{2}\left[cos\frac{\theta}{2}|0>+sin\frac{\theta}{2}e^{i\phi}|1>\right], \tag{3.17}$$

and

$$\frac{1}{2}(1-\vec{\sigma}\cdot\hat{n})|0>=sin\frac{\theta}{2}\left[sin\frac{\theta}{2}|0>-cos\frac{\theta}{2}e^{i\phi}|1>\right]. \tag{3.18}$$

The last two spinors can be easily normalized by dividing the first by $cos\frac{\theta}{2}$ and the second by $sin\frac{\theta}{2}$. This leads to the spinors

$$|\xi_n^+>=cos\frac{\theta}{2}|0>+sin\frac{\theta}{2}e^{i\phi}|1>, \tag{3.19}$$

and

$$|\xi_n^- >= sin\frac{\theta}{2}|0 > -cos\frac{\theta}{2}e^{i\phi}|1 >, \tag{3.20}$$

Since we had proved that any spinor $(1/2)(1 \pm \vec{\sigma} \cdot \hat{n})|\chi >$ is an eigenvector of $\vec{\sigma} \cdot \hat{n}$ with eigenvalues ± 1, it is obvious that the spinors $|\xi_n^+ >$ and $|\xi_n^- >$ are eigenspinors of the operator $(\vec{\sigma} \cdot \hat{n})$ with eigenvalues +1 and -1, respectively.

Exercise: Prove that the two eigenspinors $|\xi_n^+ >$ and $|\xi_n^- >$ are indeed orthogonal, i.e., in Dirac's notations,

$$< \xi_n^+ | \xi_n^- >= 0. \tag{3.21}$$

Solution

In their column format, these two spinors are given by

$$|\xi_n^+ >= cos\frac{\theta}{2}|0 > +sin\frac{\theta}{2}e^{i\phi}|1 >= \begin{bmatrix} cos\frac{\theta}{2} \\ sin\frac{\theta}{2}e^{i\phi} \end{bmatrix}, \tag{3.22}$$

which is eigenstate of $\vec{\sigma}.\vec{n}$ with eigenvalue +1, and

$$|\xi_n^- >= sin\frac{\theta}{2}|0 > -cos\frac{\theta}{2}e^{i\phi}|1 >= \begin{bmatrix} sin\frac{\theta}{2} \\ -cos\frac{\theta}{2}e^{i\phi} \end{bmatrix}, \tag{3.23}$$

which is eigenstate of $\vec{\sigma}.\vec{n}$ with eigenvalue -1. Therefore,

$$< \xi_n^+ | \xi_n^- >= cos\frac{\theta}{2}sin\frac{\theta}{2} - cos\frac{\theta}{2}sin\frac{\theta}{2} = 0 \tag{3.24}$$

which proves the result.

We note the that spinor $|\xi_n{}^- >$ can actually be obtained from the spinor $|\xi_n{}^+ >$ as follows:

$$|\xi_n{}^-(\theta, \phi) >= |\xi_n{}^+(\theta \to \pi - \theta, \phi \to \phi + \pi) > . \tag{3.25}$$

Since varying the angle θ in the range $[0, \pi]$ and the azimuthal angle ϕ in the range $[0, 2\pi]$ allows the unit vector $\hat{n}$ to sweep the entire sphere of radius 1, we can think of the spinor $|\xi_n^+ >$ as the most general form of the spinorial part of the wavefunction characterizing a spin-1/2 particle.

Hereafter, we adopt the common convention to use $|\xi_n{}^+ >$ to represent the most general form of the spinor by including an additional overall phase factor $e^{i\gamma}$ which will not affect any measurement made on the spinor,

$$|\xi_n^+ >= e^{i\gamma}\left[cos\frac{\theta}{2}|0 > +sin\frac{\theta}{2}e^{i\phi}|1 >\right]. \tag{3.26}$$

The discussion above shows that there is a one-to-one relation between the unit vector $\hat{n}$ and the spinors $|\xi_n{}^+ >$ and $|\xi_n{}^- >$ since the latter are

eigenspinors of $\vec{\sigma} \cdot \hat{n}$, which is the operator for the spin component along $\hat{n}$. This relation is represented geometrically using a sphere of radius 1 – the so-called *Bloch sphere* – where the unit vector $\hat{n}$ is the radius vector of the Bloch sphere and represents the spinor $|\xi_n{}^+>$, namely the eigenvector of $\vec{\sigma} \cdot \vec{n}$ with eigenvalue +1. This pictorial representation is very useful when analyzing the action of 2×2 unitary matrices acting on the spinor $|\xi_n{}^+>$ as actual rotations of the unit vector $\hat{n}$ on the Bloch sphere. This will be illustrated in more detail in the next sections.

3.2.2 Connection between the Bloch sphere concept and the classical interpretation of the spin of electron

For an electron characterized by the spinor $|\xi_n^+>$, it can be easily shown that the expectation values of the Pauli spin matrices are given by

$$< \xi_n^+ |\sigma_x| \xi_n{}^+ >= sin\theta cos\phi, \tag{3.27}$$

$$< \xi_n^+ |\sigma_y| \xi_n{}^+ >= sin\theta sin\phi, \tag{3.28}$$

$$< \xi_n^+ |\sigma_z| \xi_n{}^+ >= cos\theta, \tag{3.29}$$

which are exactly the Cartesian components of the unit vector $\hat{n}$.

So, we can think of the unit vector $\hat{n}$ characterized by the angles θ and ϕ as the representation of the spinor eigenstate of the operator $\vec{\sigma} \cdot \vec{n}$ with eigenvalue +1.

Similarly, starting with $|\xi_n^->$, we easily obtain

$$< \xi_n^- |\sigma_x| \xi_n{}^- >= -sin\theta cos\phi, \tag{3.30}$$

$$< \xi_n^- |\sigma_y| \xi_n{}^- >= -sin\theta sin\phi, \tag{3.31}$$

$$< \xi_n^- |\sigma_z| \xi_n{}^- >= -cos\theta, \tag{3.32}$$

which are the Cartesian coordinates of a unit vector diametrically opposite to the previous unit vector $\hat{n}$.

Therefore, any two points on the surface of the Bloch sphere which intersect a straight line going through the center can be thought as representations of eigenstates of the operator $\vec{\sigma} \cdot \hat{n}$ with eigenvalues $\pm$ 1.

Warning note:

At an early stage of exposure to the Bloch sphere concept, it is a common mistake to think of two orthogonal spinors on the Bloch sphere as unit vectors $\hat{n_1}$ and $\hat{n_2}$ whose vector product $\hat{n_1} \cdot \hat{n_2}$ must be zero, i.e., the two unit vectors $\hat{n_1}$ and $\hat{n_2}$ must be perpendicular to one another. This is NOT correct! Orthogonal spinors on the Bloch sphere do not subtend an angle of 90°; instead, they subtend an angle of 180°. This may be an alien concept that requires getting used to.

Since the spinor $|\xi_n^+>$ given explicitly in Equation (3.26) can sweep the entire Bloch sphere by varying the angles θ and ϕ, and since the expectation values of the three components of the Pauli matrices coincide with the Cartesian components of the unit vector $\hat{n}$, we can indeed think of this unit vector as representing the magnetic moment of a small magnet attached to the electron with its main axis along the direction $\hat{n}$. This interpretation is intuitively attractive but one should keep in mind that the spin of the electron is a purely quantum-mechanical concept. The unit vector $\hat{n}$ representing the general spinor $|\xi_n^+>$ on the Bloch sphere is referred to as the Bloch vector.

Since $|\xi_n^+>$ can be obtained by acting with the operator $\frac{1}{2}(1+\vec{\sigma}\cdot\hat{n})$ on an arbitrary qubit $|\chi>$, this operator is referred to as the *projection operator* in the direction $\hat{n}$. Similarly, the operator $\frac{1}{2}(1-\vec{\sigma}\cdot\hat{n})$ is the *projection operator* in the direction $-\hat{n}$.

Exercise: Prove that if a spinor is in the state $|\xi_{n_1}^+>$, its components along the two spinors $|\xi_{n_2}^+>$, $|\xi_{n_2}^->$, which are eigenstates of the projection operator along another unit vector $\hat{n_2}$, are such that

$$|<\xi_{n_2}^+|\xi_{n_1}^+>|^2+|<\xi_{n_2}^-|\xi_{n_1}^+>|^2=1. \tag{3.33}$$

This is easy to prove and is left for the reader.

3.2.3 Relationship with qubit

The spinor $|\xi_n^+>$ can be easily related to a qubit $|\chi>$ defined as a column vector in $\mathcal{C}^2$:

$$|\chi>=\begin{bmatrix}\alpha\\ \beta\end{bmatrix}, \tag{3.34}$$

where α and β are complex numbers. This last expression can be rewritten as

$$|\chi>=\begin{bmatrix}Re(\alpha)+iIm(\alpha)\\ Re(\beta)+iIm(\beta)\end{bmatrix}, \tag{3.35}$$

or equivalently,

$$|\chi>=\begin{bmatrix}|\alpha|e^{i\phi_\alpha}\\ |\beta|e^{i\phi_\beta}\end{bmatrix}. \tag{3.36}$$

Hence,

$$|\chi>=e^{i\phi_\alpha}\begin{bmatrix}|\alpha|\\ |\beta|e^{i(\phi_\beta-\phi_\alpha)}\end{bmatrix}, \tag{3.37}$$

which has the general form of $|\xi_n^+>$ in Equation (3.26) for the following choice of parameters γ, θ and ϕ:

$$\gamma=\phi_\alpha, \tag{3.38}$$

$$\theta=2arctan\left(\frac{\sqrt{1-|\alpha|^2}}{|\alpha|}\right), \tag{3.39}$$

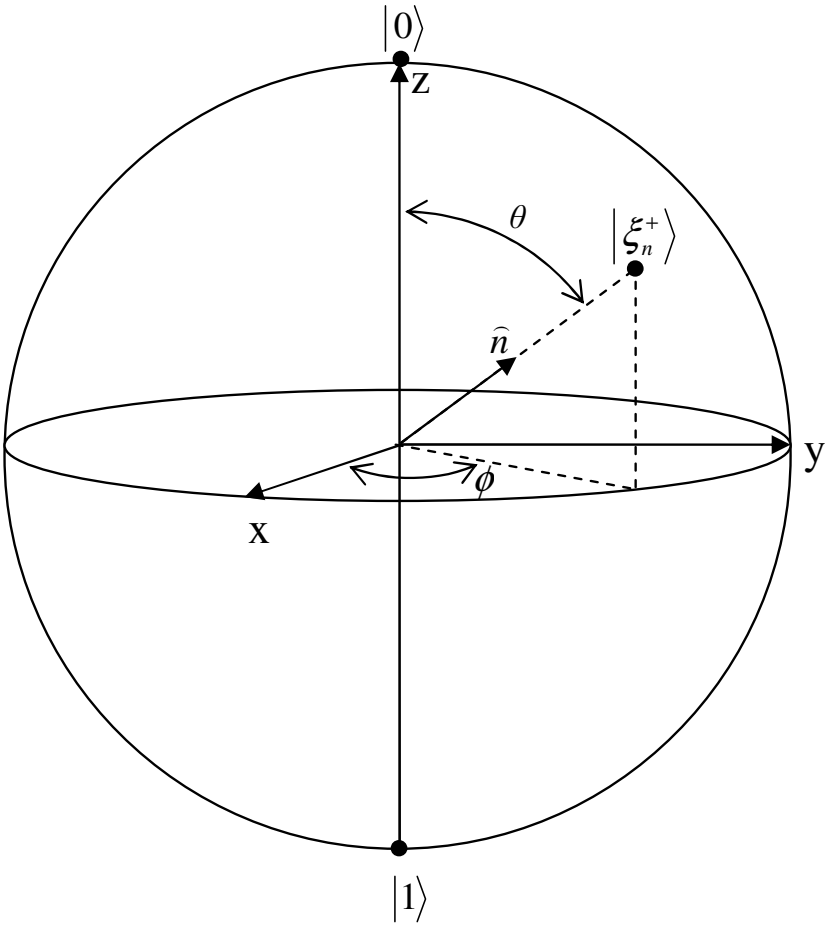


FIGURE 3.1
Bloch sphere representation of a qubit.

and

$$\phi = \phi_\beta - \phi_\alpha. \tag{3.40}$$

3.2.4 Special spinors

3.2.4.1 North and South Poles

Referring to Equation (3.26), it is obvious that the state $|0>$ can be written as

$$|0> = |\xi_n^+(\theta = 0, \phi, \gamma)>, \tag{3.41}$$

and the state $|1>$ as

$$|1> = |\xi_n^+(\theta = \pi, \phi, \gamma)> . \tag{3.42}$$

The fact that ϕ and γ are arbitrary is irrelevant since observables should not depend on the overall phase of the spinor. This interpretation is consistent with associating the spinor $|0>$ with the point representing the north pole on the Bloch sphere whereas the spinor $|1>$, which is orthogonal to $|0>$, is associated with the south pole on the Bloch sphere. Since they are orthogonal, they are diametrically opposite through the sphere center.

3.2.4.2 Reaching the equator

The Hadamard matrix is one of the most widely used matrix in the design of quantum logic gates and is given by

$$H = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}. \tag{3.43}$$

One of the two eigenvectors of the Hadamard matrix is given by

$$|\chi>_{Hadamard} = cos\left(\frac{\pi}{8}\right)|0> + sin\left(\frac{\pi}{8}\right)|1> . \tag{3.44}$$

We wish to find the locations on the Bloch sphere of this eigenvector of the Hadamard matrix. If we express this eigenvector as $\alpha|0> + \beta|1>$, then following the usual prescription:

$$\alpha = e^{i\gamma} cos\frac{\theta}{2} = cos\left(\frac{\pi}{8}\right), \tag{3.45}$$

$$\beta = e^{i\gamma} e^{i\phi} sin\frac{\theta}{2} = sin\left(\frac{\pi}{8}\right). \tag{3.46}$$

Hence, $\theta = \frac{\pi}{4}$, $\phi = 0$ and γ can be any integral multiple of 2π.

Hence the eigenvector of the Hadamard matrix given by Equation (3.44) is in the x-z plane, halfway between the north pole and the point where the positive x-axis pierces the Bloch sphere.

It is then easy to show that the other eigenvector of the Hadamard matrix is located diametrically opposite to the unit vector associated with the spinor in Equation (3.44).

3.2.5 The spin flip matrix

We now address the following question: What is the 2×2 matrix M which changes $|\xi_n^- >$ into $|\xi_n^+ >$?, i.e.,

$$|\xi_n^+ >= e^{i\gamma} \begin{bmatrix} cos\frac{\theta}{2} \\ sin\frac{\theta}{2} e^{i\phi} \end{bmatrix} = M|\xi_n^- >= Me^{i\gamma} \begin{bmatrix} sin\frac{\theta}{2} \\ -cos\frac{\theta}{2} e^{i\phi} \end{bmatrix}. \tag{3.47}$$

Evidently

$$M = \begin{bmatrix} 0 & -e^{-i\phi} \\ e^{i\phi} & 0 \end{bmatrix}, \tag{3.48}$$

and indeed, we find

$$M|\xi_n^- >= \begin{bmatrix} cos\frac{\theta}{2} \\ sin\frac{\theta}{2} e^{i\phi} \end{bmatrix}, \tag{3.49}$$

as required.

M can be written as a product of some well-known 2×2 matrices

$$M = \begin{bmatrix} 1 & 0 \\ 0 & e^{i\phi} \end{bmatrix} \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & e^{-i\phi} \end{bmatrix} \tag{3.50}$$

or, in a more compact form

$$M = e^{i\frac{\pi}{2}} P(\phi)\sigma_y P(-\phi), \tag{3.51}$$

where $P(\phi)$ is the phase shift matrix (see Problem 3.4).

3.2.6 Excursions on the Bloch sphere: the Pauli matrices revisited

In Chapter 2, we proved the identity

$$e^{i\theta A} = cos\theta I + isin\theta A, \tag{3.52}$$

where θ is real and the matrix A is such that $A^2 = I$. This can be looked at as a generalization to operators of the well-known Euler relation for complex numbers, i.e., $e^{iz} = cosz + isinz$.

We are now going to illustrate the usefulness of this relation for finding the unitary matrices which allow us to perform rotations on the Bloch sphere around the three axes x, y and z. Since any general spinor in $\mathcal{C}^2$ can be written as a linear superposition of the spinors $|0 >$ and $|1 >$ (see Equation (3.1), we start with each of these two spinors and see how they are modified after being rotated through some angle θ as indicated in Figure 3.2.

Consider first the spinor $|0 >$ and let us seek the matrix which would rotate that spinor an angle θ away from the z-axis as indicated in Figure 3.2. After rotation in the (y,z) plane (anticlockwise rotation, i.e., pushing the y-axis

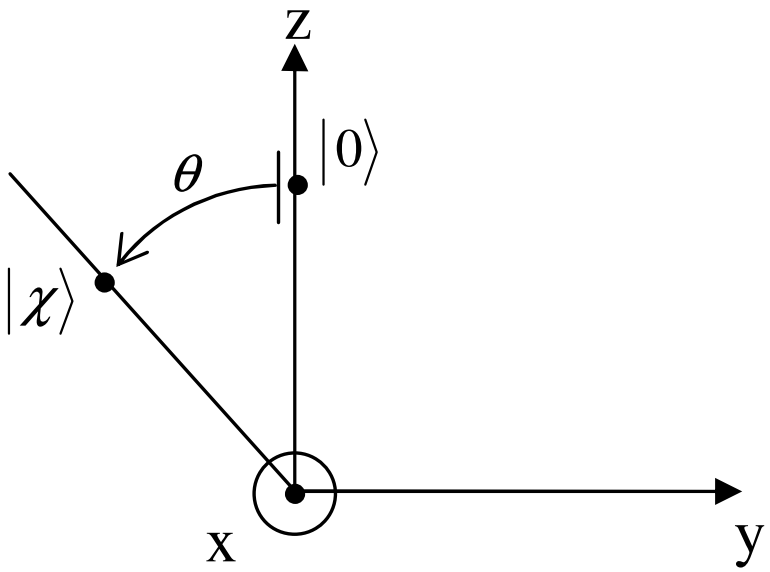


FIGURE 3.2
Illustration of the rotation of the spinor $|0>$ to the spinor $|\chi>$. The rotation is counterclockwise around the x-axis.

towards the z-axis using the shortest path), the new spinor $|\chi>_{new}$ is given by

$$|\chi>_{new}= cos\frac{\theta}{2}|0> +sin\frac{\theta}{2}e^{-i\frac{\pi}{2}}|1>, \tag{3.53}$$

or

$$|\chi>_{new}= \begin{bmatrix} cos\frac{\theta}{2} \\ sin\frac{\theta}{2}e^{-i\pi/2} \end{bmatrix} = \begin{bmatrix} cos\frac{\theta}{2} \\ -isin\frac{\theta}{2} \end{bmatrix}. \tag{3.54}$$

Since the rotation matrix is such that

$$|\chi>_{new}= R_x(\theta)|0>, \tag{3.55}$$

the spinor in the last equation is the first column of the matrix $R_x(\theta)$.

Next, we apply $R_x(\theta)$ to the spinor $|1>$ and find (see Figure 3.3)

$$|\chi^{'}>_{new}= R_x(\theta)|1>= cos(\frac{\pi-\theta}{2})|0> +e^{i\pi/2}sin(\frac{\pi-\theta}{2})|1>, \tag{3.56}$$

which has the explicit form

$$|\chi^{'}>= \begin{bmatrix} sin\frac{\theta}{2} \\ icos\frac{\theta}{2} \end{bmatrix} = i\begin{bmatrix} -isin\frac{\theta}{2} \\ cos\frac{\theta}{2} \end{bmatrix}. \tag{3.57}$$

This second spinor represents the second column vector of the matrix $R_x(\theta)$. However, since the rotation matrix must leave the norm of the spinor unchanged, it must be a 2 × 2 unitary matrix. If we drop the overall phase factor i of the last column vector in Equation (3.57), an expression for the 2 × 2 matrix describing a rotation around the x-axis of a general qubit (which is a superposition, i.e., linear combination of the $|0>$ and $|1>$ bits) is given by

$$R_x(\theta) = \begin{bmatrix} cos\frac{\theta}{2} & -isin\frac{\theta}{2} \\ -isin\frac{\theta}{2} & cos\frac{\theta}{2} \end{bmatrix}, \tag{3.58}$$

which is unitary as easily checked.

Using the identity (3.52), the matrix $R_x(\theta)$ can be generated from the Pauli matrix σ_x as follows

$$R_x(\theta) = e^{-i\frac{\theta}{2}\sigma_x}. \tag{3.59}$$

Using similar arguments, it can be shown that for rotation of a general spinor around the y-axis on the Bloch sphere, the rotation matrix is

$$R_y(\theta) = \begin{bmatrix} cos\frac{\theta}{2} & -sin\frac{\theta}{2} \\ sin\frac{\theta}{2} & cos\frac{\theta}{2} \end{bmatrix}, \tag{3.60}$$

and, for a rotation around the z-axis,

$$R_z(\theta) = \begin{bmatrix} e^{-i\frac{\theta}{2}} & 0 \\ 0 & e^{i\frac{\theta}{2}} \end{bmatrix}. \tag{3.61}$$

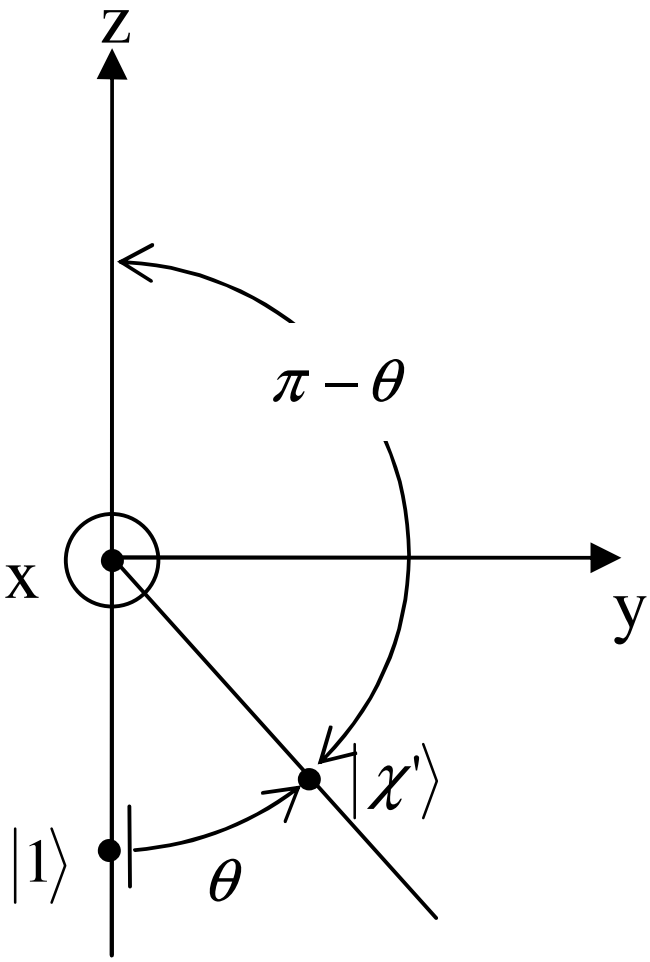


FIGURE 3.3
Illustration of the rotation of the spinor $|1>$ to the spinor $|\chi^{'}>$ to find the second portion of the rotation matrix of a general spinor around the x-axis.

With the use of the identity (3.52), these last relations can be recast as follows

$$R_y(\theta) = e^{-i\frac{\theta}{2}\sigma_y}, \tag{3.62}$$

and

$$R_z(\theta) = e^{-i\frac{\theta}{2}\sigma_z}. \tag{3.63}$$

Because the Pauli matrices constitute the building blocks to describe rotations on the Bloch sphere with the x-, y- or z-axis as the axis of rotations, they are sometimes referred to as the generators of the rotations in the Hilbert space of spinors associated with spin-1/2 particles [2].

Some sophisticated studies of the evolution of spinors on the Bloch sphere can be found in references [5, 6, 7]. The interested reader is referred to these works for further reading.

3.3 Problems

- **Problem 3.1**

  Consider the following matrix

  $$\begin{pmatrix} 1 & 3 \\ 3 & 1 \end{pmatrix}. \tag{3.64}$$

  Calculate the eigenvalues and corresponding eigenvectors of this matrix. What are the angles θ and ϕ in the general expression of the spinor on the Bloch sphere, i.e.,

  $$|\xi_n^+ >= cos\frac{\theta}{2}|0 > +sin\frac{\theta}{2}e^{i\phi}|1 >, \tag{3.65}$$

  associated with the two eigenvectors of the above matrix?

- **Problem 3.2**

  Consider the following qubit

  $$|\chi >= \sqrt{\frac{3}{5}}|0 > -\sqrt{\frac{2}{5}}|1 > . \tag{3.66}$$

  Calculate the angles γ, θ and ϕ associated with the corresponding spinor and show its location on the Bloch sphere.

- **Problem 3.3**

  Prove that Equation (3.44) is indeed the eigenvector of the Hadamard matrix H with eigenvalue $+1$. Find the other eigenvector of the Hadamard matrix and its corresponding eigenvalue.

- **Problem 3.4**

  Locate the points on the Bloch sphere which represent the eigenvectors of the Pauli matrices (σ_x, σ_y, σ_z).

  Two other matrices frequently used in the design of quantum logic gates are the S matrix

$$S = \begin{pmatrix} 1 & 0 \\ 0 & i \end{pmatrix}, \tag{3.67}$$

  which is a special case of the phase shift matrix P

$$P(\phi) = \begin{pmatrix} 1 & 0 \\ 0 & e^{i\phi} \end{pmatrix}, \tag{3.68}$$

  and the $\pi/8$ or T-matrix

$$T = \begin{pmatrix} 1 & 0 \\ 0 & e^{i\pi/4} \end{pmatrix}. \tag{3.69}$$

  Calculate the eigenspinors and eigenvalues of the S and T matrices. Locate the corresponding eigenvectors on the Bloch sphere.

- **Problem 3.5**

  Show that the spin flip matrix can also be written as a product of some phase shift matrices and the Pauli matrix σ_x.

- **Problem 3.6**

  The operator $\frac{1}{2}(1 + \vec{\sigma} \cdot \hat{n})$ is referred to as a *projection operator*. When applied to an arbitrary spinor $|\chi >$, the resulting state vector is an eigenstate of the operator $(\vec{\sigma} \cdot \hat{n})$, i.e., a spinor whose representation on the Bloch sphere is given by the unit vector $\hat{n}$. Using $|\chi >$ equal to $|0 >$, $|1 >$, and $\frac{1}{\sqrt{2}}(|0 > + |1 >)$, show that the three corresponding spinors $\frac{1}{2}(1+\vec{\sigma}\cdot\hat{n})|\chi >$ may differ by a overall phase factor but are all associated with a point on the Bloch sphere characterized by the Bloch vector $\hat{n}$.

  The projection operators $\frac{1}{2}(1 \pm \vec{\sigma} \cdot \hat{n})$ are special cases of a very general decomposition of any 2×2 matrix in terms of the 2×2 identity and three Pauli matrices discussed in Chapter 2, i.e.,

$$M = \begin{pmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{pmatrix} \tag{3.70}$$

  can be decomposed as follows

$$M = \frac{m_{11} + m_{22}}{2} I + \frac{m_{11} - m_{22}}{2}\sigma_z + \frac{m_{12} + m_{21}}{2}\sigma_x + \frac{m_{12} - m_{21}}{2}\sigma_y, \tag{3.71}$$

which can be written in a more condensed form

$$M = a_o I + \vec{a} \cdot \vec{\sigma}, \tag{3.72}$$

where

$$a_o = \frac{1}{2} Tr(M), \tag{3.73}$$

and

$$\vec{a} = \frac{1}{2} Tr(M\vec{\sigma}), \tag{3.74}$$

where Tr is the trace operator and $\vec{\sigma} = \sigma_x \hat{x} + \sigma_y \hat{y} + \sigma_z \hat{z}$.

Find the value of a_0 and $\vec{a}$ for the matrices H, T and S defined in Problem 3.4.

- **Problem 3.7**

  Consider the two matrices M_1 and M_2 with the decomposition

$$M_1 = {a_o}^{(1)} I + \vec{a_1}.\vec{\sigma}, \tag{3.75}$$

and

$$M_2 = {a_o}^{(2)} I + \vec{a_2}.\vec{\sigma}. \tag{3.76}$$

  What must be the relationship between the two vectors $\vec{a_1}$ and $\vec{a_2}$ for the matrices M_1 and M_2 to commute?

  Hint: Use the identity

$$(\vec{\sigma} \cdot \vec{a_1})(\vec{\sigma} \cdot \vec{a_2}) = (\vec{a_1} \cdot \vec{a_2})I + i\vec{\sigma} \cdot (\vec{a_1} \times \vec{a_2}), \tag{3.77}$$

  where I is the 2×2 identity matrix.

- **Problem 3.8**

  Prove the following identity for the Pauli spin matrices:

$$\sigma_x \sigma_y \sigma_z = iI. \tag{3.78}$$

  Using the Bloch sphere concept, interpret this relation geometrically.

- **Problem 3.9**

  Generalize the previous problem to show that for any set of three vectors $(\vec{a}, \vec{b}, \vec{c})$ forming a right hand orthogonal basis, the following relation is true.

$$(\vec{\sigma} \cdot \vec{a})(\vec{\sigma} \cdot \vec{b})(\vec{\sigma} \cdot \vec{c}) = iI. \tag{3.79}$$

  Hint: Use the identity

$$(\sigma \cdot \vec{a})(\sigma \cdot \vec{b}) = \vec{a} \cdot \vec{b} I + i\vec{\sigma} \cdot (\vec{a} \times \vec{b}). \tag{3.80}$$

- **Problem 3.10**

  Prove that the following is true

  $$R_y(\theta_1)R_y(\theta_2) = R_y(\theta_2)R_y(\theta_1) = R_y(\theta_1 + \theta_2), \tag{3.81}$$

  i.e., rotations about a given axis are commutable. Prove similar relations with R_y replaced by R_x and R_z.

  Hint: This can be shown easily by making use of the following trigonometric relations

  $$sin(\alpha \pm \beta) = sin\alpha cos\beta \pm cos\alpha sin\beta, \tag{3.82}$$

  and

  $$cos(\alpha \pm \beta) = cos\alpha cos\beta \mp sin\alpha sin\beta. \tag{3.83}$$

- **Problem 3.11**

  Show that for any orthogonal basis $(\vec{a}, \vec{b}, \vec{c})$, the following operator will bring any spinor back to its original location on the Bloch sphere (with an additional phase shift). What is the expression for that phase shift?

  $$R_{\vec{c}}(\pi)R_{\vec{b}}(\pi)R_{\vec{a}}(\pi). \tag{3.84}$$

  Any cyclic permutation of the three vectors leads to a similar result.

- **Problem 3.12**

  Equation (3.51) gives an expression for the *spin flip matrix* in terms of the phase shift matrix and the Pauli matrix σ_y. Interpret this equation geometrically as a succession of rotations on the Bloch sphere keeping in mind that the Pauli matrices were identified as the generators of spinor rotations on the Bloch sphere.

- **Problem 3.13**

  Using the Bloch sphere concept, describe geometrically the action of the matrices S and T on a general Bloch vector $\hat{n}$ or qubit $|\xi_n^+>$.

- **Problem 3.14**

  What is the 2×2 matrix allowing transition from point 1 to 2, 2 to 3 and 3 to 1 on the Bloch sphere as shown in Figure 3.4?

- **Problem 3.15**

  Using the fact that

  $$[\sigma_x, \sigma_y] = 2i\sigma_z, \tag{3.85}$$

  calculate the commutator $[R_x(\theta_1), R_y(\theta_2)]$ where the rotation matrices are given by

  $$R_x(\theta_1) = e^{-i\frac{\theta_1}{2}\sigma_x}, \tag{3.86}$$

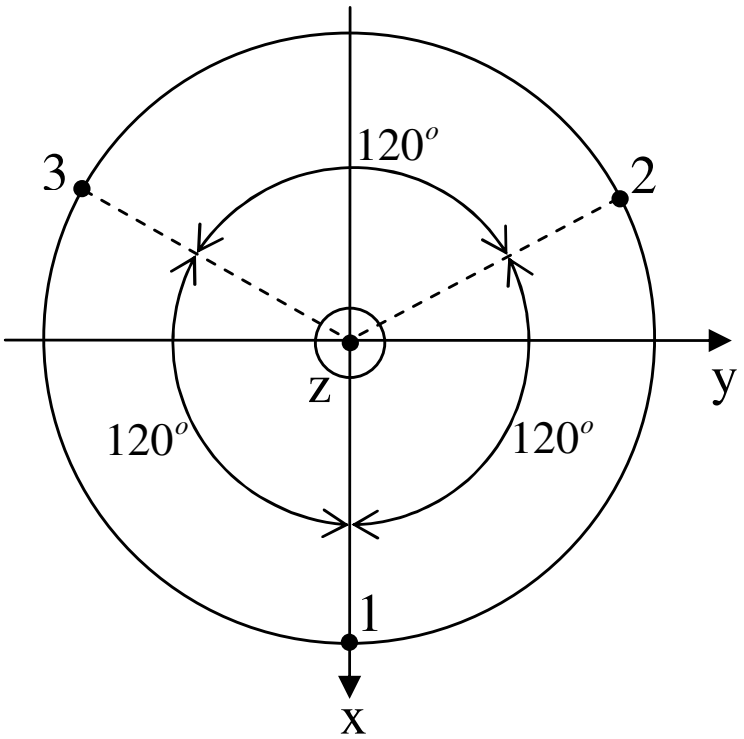


FIGURE 3.4
Locations of three spinors (1-2-3) located on the circle at the intersection of the Bloch sphere and the (x-y) plane.

and

$$R_y(\theta_2) = e^{-i\frac{\theta_2}{2}\sigma_y}. \tag{3.87}$$

Hint: Use Equation (3.52).

- **Problem 3.16**

  Consider the "Square-Root-of-Not" matrix defined as [3]

$$\sqrt{NOT} = \begin{pmatrix} \frac{1+i}{2} & \frac{1-i}{2} \\ \frac{1-i}{2} & \frac{1+i}{2} \end{pmatrix}. \tag{3.88}$$

  (1) Show that this matrix is unitary.

  (2) Find the values of a_o and $\vec{a}$ in the decomposition of the $\sqrt{NOT}$ matrix in the form

$$M = a_o I + \vec{a} \cdot \vec{\sigma} \tag{3.89}$$

  discussed in a previous exercise.

  (3) Show that the product of the $\sqrt{NOT}$ matrix with itself is σ_x, the bit flip matrix. (Hence $\sqrt{NOT} \cdot \sqrt{NOT} = NOT$, where NOT is the bit flip operation performed by a NOT-gate).

  (4) Is the matrix $\sqrt{NOT}$ defined above unique?

  (5) Calculate the eigenvalues and eigenvectors of the $\sqrt{NOT}$ matrix and show the locations of its corresponding eigenvectors on the Bloch sphere.

- **Problem 3.17**

  It can be shown that any 2×2 unitary operator can be written in the form [4]

$$U = e^{i\alpha} R_{\hat{n}}(\theta), \tag{3.90}$$

  for some real numbers α and θ and a real three dimensional vector $\hat{n}$.

  Find the values of α, θ and $\hat{n}$, for the three Pauli matrices, Hadamard, Phase Shift, and $\pi/8$ gates.

3.4 References

[1] S-L Zhu, Z.D. Wang and Y-D Zhang, "Nonadiabatic noncyclic geometric phase of a spin-1/2 particle subject to an arbitrary magnetic field", Phys. Rev. B, **61**, 1142 (2000).

[2] A more formal discussion of the Pauli matrices, including their use as generators of infinitesimal rotations on spinors and their generalization to the Dirac matrices used in relativistic quantum mechanics can be found in Chapter 4 of G. Arfken, *Mathematical Methods for Physicists*, 2^{nd} Edition, (Academic Press, New York, 1970).

[3] C. P. Williams and S. H. Clearwater, *Exploration in Quantum Computing*, (Springer-Verlag, 1998). (Telos, the Electronic Library of Science).

[4] M.A. Nielsen and I.L. Chuang, *Quantum Computation and Quantum Information*, (Cambridge University Press, New York, 2000).

[5] N.D. Mermin, "From Cbits to Qbits: Teaching computer scientists quantum mechanics", Am. J. Phys., **71**, 23 (2002).

[6] D. M. Tong, J-L Chen, L.C. Kwek, C.H. Lai and C.H. Oh, "General formalism of Hamiltonians for realizing a prescribed evolution of a qubit", Phys. Rev. A, **68**, 062307 (2003).

[7] D.M. Sullivan and D.S. Citrin, "Time-domain simulation of quantum spin", J. Appl. Phys., **94**, 6518 (2003).

4

Evolution of a Spinor on the Bloch Sphere

In this chapter, we will consider the general approach used to calculate the time evolution of the probability of a spinor reaching a particular location on the Bloch sphere from another location, when subjected to the simultaneous actions of two spatially uniform magnetic fields. One of these fields is static (time-independent) and directed along a specific axis, while the other is time-dependent, has a different amplitude, and rotates in the plane perpendicular to the first magnetic field. This problem was originally considered by I. I. Rabi who calculated the probability of a "spin flip", i.e., the probability for a spinor, originally located at the north pole on the Bloch sphere, to ultimately reach the south pole [1, 2, 3]. Rabi calculated the time evolution of the probability in the case of a constant (dc) magnetic field along the z-axis and a circularly polarized time-varying (ac) magnetic field of different amplitude, with a specific frequency, rotating in the (x-y) plane. This led to the derivation of what is known as the *Rabi formula* which is one of the key ingredients in Nuclear Magnetic Resonance (NMR). In most textbooks, the Rabi formula is derived by solving the time-dependent Pauli Equation [Equation (2.31)] with a judicious choice of the reference frame to describe the time-evolution of the 2-component spinor wavefunction. Here, we use a different approach based on the concept of the Bloch sphere described in the previous chapter. This approach is more instructive since it directly relates to NMR concepts and is more synergistic with concepts used in spin-based quantum computing.

4.1 Spin-1/2 particle in a constant magnetic field: Larmor precession

We use the well-known Ehrenfest theorem of quantum mechanics (see Chapter 15) to calculate the time evolution of the spin of an electron in a spatially uniform but time-dependent magnetic field $\vec{B}(t)$. Specifically, we will consider an electron's spin in a bulk piece of semiconductor for which (see Problem 4.1) the time evolution of the (spatial) average values of the Pauli spin matrices is governed by the following set of coupled equations

$$\frac{d}{dt} < \sigma_x > = \frac{g\mu_B}{\hbar}(B_y < \sigma_z > - B_z < \sigma_y >)$$
$$\frac{d}{dt} < \sigma_y > = \frac{g\mu_B}{\hbar}(B_z < \sigma_x > - B_x < \sigma_z >)$$
$$\frac{d}{dt} < \sigma_z > = \frac{g\mu_B}{\hbar}(B_x < \sigma_y > - B_y < \sigma_x >), \tag{4.1}$$

where the brackets $< ... >$ denote an average of the operators over the ket $|\xi_n^+ >$ representing the spinor on the Bloch sphere, i.e., $< \sigma_n >=< \xi_n^+ | \sigma_n | \xi_n^+ >$, and g is the effective Landé g-factor in the material.

Equation (4.1) can be written in a matrix form

$$\frac{d}{dt}\begin{bmatrix} < \sigma_x > \\ < \sigma_y > \\ < \sigma_z > \end{bmatrix} = \frac{g\mu_B}{\hbar}\begin{bmatrix} 0 & -B_z & B_y \\ B_z & 0 & -B_x \\ -B_y & B_x & 0 \end{bmatrix}\begin{bmatrix} < \sigma_x > \\ < \sigma_y > \\ < \sigma_z > \end{bmatrix}, \tag{4.2}$$

or, more compactly, in the vectorial form

$$\frac{d < \vec{\sigma} >}{dt} = \frac{g\mu_B}{\hbar}(\vec{B} \times < \vec{\sigma} >) = \vec{\Omega} \times < \vec{\sigma} >, \tag{4.3}$$

where the vector $< \vec{\sigma} > = (< \sigma_x >, < \sigma_y >, < \sigma_z >)$ and $\vec{\Omega} = \frac{g\mu_B}{\hbar}\vec{B}$, which is also equal to $\frac{e\vec{B}}{m_0}$ for a free electron since $g_{free-electron} = 2$ and $\mu_B = e\hbar/(2m_0)$. This is the well known Larmor precession frequency. If the magnetic field is time-independent, then the Larmor precession frequency is also time-independent; otherwise, it follows the time dependence of the magnetic field.

Because the operator for any component of the spin angular momentum along a coordinate axis is given by $S_n = (\hbar/2)\sigma_n$, it is obvious that Equation (4.3) can be recast as

$$\frac{d < \vec{S} >}{dt} = \frac{g\mu_B}{\hbar}(\vec{B} \times < \vec{S} >) = \vec{\Omega} \times < \vec{S} >, \tag{4.4}$$

where $< \vec{S} >$ is, by definition, the expected value of the spin angular momentum. Equation (4.4) is the familiar Larmor equation describing Larmor precession of a spin about a magnetic field.

Exercise: Consider the case of a spatially uniform and time-independent magnetic field along the z-axis. For that case, prove that the spin angular momentum vector $\vec{S}$ rotates around the z-axis with a constant angular frequency and find an expression for that angular frequency.

Solution

Equation (4.4) leads to

$$\vec{B} \cdot \frac{d\vec{S}}{dt} = 0. \tag{4.5}$$

Since $\vec{B}$ is time-independent, we obtain that $\frac{d(\vec{S}\cdot\vec{B})}{dt} = 0$, i.e., the angle θ between the two vectors $\vec{B}$ and $\vec{S}$ is time independent. In other words, the vector $\vec{S}$ moves on a cone with a fixed apex angle 2θ, and the axis of the cone coincides with the magnetic field $\vec{B}$.

Using spherical coordinates and taking into account the fact that the polar angle θ is time independent, Equation (4.4) immediately shows that the azimuthal angle ϕ (which the angle that the projection of $\vec{S}$ on the (x-y) plane subtends with the x-axis) changes in time with a constant rate

$$\frac{d\phi}{dt} = \frac{g\mu_B B_z}{\hbar}. \tag{4.6}$$

Therefore, the spin angular momentum vector $\vec{S}$ rotates on a cone which makes a fixed angle with the magnetic field $\vec{B}$, and the rotation takes place at a constant angular frequency equal to the Larmor precession frequency.

4.1.1 Rotation on the Bloch sphere

Suppose we have a spinor initially located at the north pole on the Bloch sphere and we would like to flip it to the south pole using some combination of magnetic fields. What is the simplest combination of magnetic fields required to effect this transition?

In the previous section, we derived the equation of motion of a spinor in an arbitrary (including time-varying) spatially uniform magnetic field, and showed that the spinor will precess about the field in a cone with an angular frequency that is proportional to the strength of the magnetic field. So, if all we have is a constant magnetic field along the z-axis, nothing will happen to the spinor originally located at the north pole; it will remain at the north pole (polar angle $\theta = 0$). But what if we apply a magnetic field in the equatorial plane of the Bloch sphere (i.e., the x-y plane) to force the spinor to move away from the north pole? If this field acts alone, the cone on which the spinor will rotate will actually be a circle on the Bloch sphere going through the north and south poles. If that new magnetic field is selected to be along the x-axis, the circle would be located in the (y-z) plane. This does take the spinor to the south pole periodically and effect periodic spins flips, but only if the new field B_{new} acts alone and the original field B_{old} along the z-axis is removed.

If we leave B_{old} on, then as soon as the spinor starts to rotate in the (y-z) plane under the influence of B_{new}, B_{old} will try to make it move in a cone around the z-axis. The spinor will then come out of the (y-z) plane and the only way B_{new} can keep on exerting the right torque on the spinor to eventually bring it down to the south pole is to *chase* the spinor so that B_{new} always stays perpendicular to the direction of the spinor $\hat{n}$. This means that B_{new} needs to rotate in the equatorial plane at the right angular frequency to keep up with the spinor as it is trying to rotate on a cone around the z-axis. This situation is shown in Fig. 4.1. The magnetic field B_{new} in the equatorial

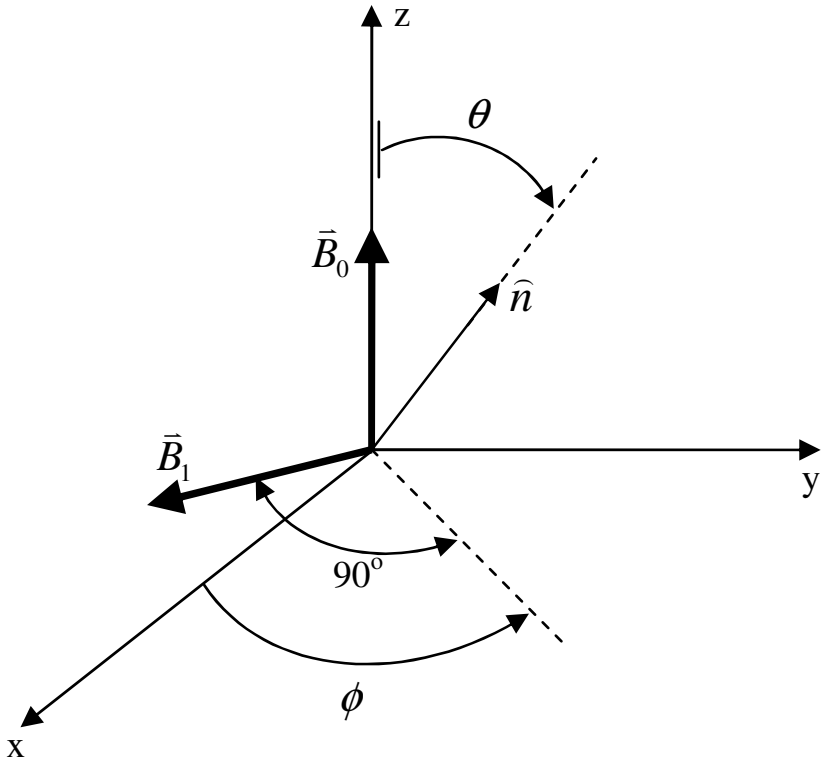


FIGURE 4.1
Spin flipping between north and south poles is achieved via the combination of a spatially uniform time-independent magnetic field along the z-axis and a rotating magnetic field in the (x-y) plane ***chasing*** the unit vector $\hat{n}$ whose Larmor precession is controlled by the uniform magnetic field along the z-axis. Here $B_1 = B_{new}$ and $B_0 = B_{old}$.

plane should therefore be a constant magnetic field rotating on the equatorial plane at the angular frequency given by Equation (4.6).

How fast the spinor will be able to reach the south pole will depend on the strength of B_{new}. As B_{new} chases the spinor $\hat{n}$, the latter rotates in a circle on the plane defined by the z-axis and $\hat{n}$, with an angular frequency proportional to the strength of B_{new}. With both B_{old} and B_{new} present, the spinor $\hat{n}$ will periodically visit the north and south poles with a period inversely proportional to the strength of B_{new}. This is the basic physics underlying nuclear magnetic resonance (NMR) [4] which is routinely used in numerous applications and more recently for practical implementations of quantum computers [5].

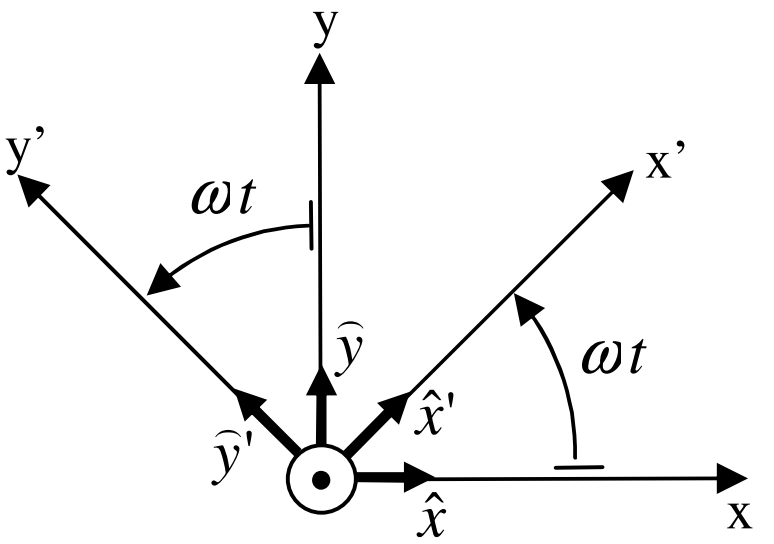


FIGURE 4.2
Illustration of the two reference frames used to described the temporal evolution of a spinor under the combined actions of B_{new} and B_{old}. The reference frame $(x^{'}, y^{'}, z^{'})$ rotates counterclockwise around the z-axis at a constant angular frequency ω.

An exact calculation of the probability of spin flip via the combined actions of B_{old} and B_{new} was originally performed by I. I. Rabi in the 1940s but here we present a different derivation based on the Bloch sphere concept discussed in the previous chapter and the results of the previous section.

4.2 Preparing to derive the Rabi formula

Consider a fixed (time-independent) reference frame with Cartesian coordinates (x, y, z) and a second time-varying reference frame with Cartesian coordinates ($x^{'}(t)$, $y^{'}(t)$, $z^{'}(t)$) obtained from the initial reference frame through an anti-clockwise rotation around the z-axis at an angular frequency ω. This is shown in Figure 4.2.

Referring to Figure (4.2), it is easy to see that the relationship between the unit vector in the unprimed and primed reference frames is given by

$$\hat{x^{'}} = cos\omega t\hat{x} + sin\omega t\hat{y}$$

$$\hat{y'} = -sin\omega t\hat{x} + cos\omega t\hat{y}$$
$$\hat{z'} = \hat{z}. \tag{4.7}$$

Using these relations, any three-dimensional vector $\hat{n}$ can be decomposed in the two reference frames as follows

$$\hat{n} = n_{x'}\hat{x'} + n_{y'}\hat{y'} + n_{z'}\hat{z'} = n_x\hat{x} + n_y\hat{y} + n_z\hat{z}, \tag{4.8}$$

where (n_x, n_y, n_z) and $(n_{x'}, n_{y'}, n_{z'})$ are the components in the unprimed and primed reference frames, respectively. These components are related as follows

$$\hat{n'} = \begin{bmatrix} n_{x'} \\ n_{y'} \\ n_{z'} \end{bmatrix} = [A] \begin{bmatrix} n_x \\ n_y \\ n_z \end{bmatrix} = [A]\hat{n}, \tag{4.9}$$

where $[A]$ is given by

$$[A] = \begin{bmatrix} cos\omega t & sin\omega t & 0 \\ -sin\omega t & cos\omega t & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{4.10}$$

Exercise: Prove that the transformation matrix $[A]$ is indeed given by Equation(4.10)

The reader can easily prove this result using Equations (4.7) - (4.9).

Earlier, we saw that when a spinor is subjected to a spatially uniform time dependent magnetic field, its time evolution is given by Equation (4.2). Therefore, in the unprimed reference frame, we have

$$\frac{d\hat{n}}{dt} = [X]\hat{n}, \tag{4.11}$$

where the matrix $[X]$ is given explicitly by the 3×3 matrix in Equation (4.2).

Next, we ask what is the analytical form of the matrix $[X']$ such that

$$\frac{d\hat{n}'}{dt} = [X']\hat{n}', \tag{4.12}$$

which describes the time evolution of the spinor in the rotating frame.

From Equation (4.9), we get

$$\frac{d\hat{n}}{dt} = [A]^{-1}\frac{d\hat{n'}}{dt} + \frac{d[A]^{-1}}{dt}\hat{n'}, \tag{4.13}$$

where (invert Equation (4.10))

$$[A]^{-1} = \begin{bmatrix} cos\omega t & -sin\omega t & 0 \\ sin\omega t & cos\omega t & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{4.14}$$

Hence,

$$[A]^{-1}\frac{d\hat{n'}}{dt} + \frac{d[A]^{-1}}{dt}\hat{n'} = [X]\hat{n} = [X][A]^{-1}\hat{n'}, \tag{4.15}$$

which can be rearranged as follows

$$\frac{d\hat{n'}}{dt} = [A]\left\{[X][A]^{-1} - \frac{d[A]^{-1}}{dt}\right\}\hat{n'} = \left\{[A][X][A]^{-1} - [A]\frac{d[A]^{-1}}{dt}\right\}\hat{n'} \tag{4.16}$$

where the expression between the last curly brackets is the matrix $[X']$.

Exercise: If a spin-1/2 particle is subjected to the simultaneous actions of B_{new} and B_{old} (see Figure 4.1) then the total magnetic field it experiences is given by

$$\vec{B} = (B_{new}cos\omega t, B_{new}sin\omega t, B_{old}). \tag{4.17}$$

Show that the matrix $[X']$ is

$$[X'] = \begin{bmatrix} 0 & \omega - \omega_0 & 0 \\ -(\omega - \omega_0) & 0 & -\omega_1 \\ 0 & \omega_1 & 0 \end{bmatrix}, \tag{4.18}$$

where

$$\omega_0 = \frac{g\mu_B B_{old}}{\hbar}, \tag{4.19}$$

and

$$\omega_1 = \frac{g\mu_B B_{new}}{\hbar}. \tag{4.20}$$

Note that ω_0 is the Larmor frequency associated with the z-directed static magnetic field and ω_1 is the Larmor frequency associated with the time varying magnetic field rotating in the x-y plane.

Solution

Use the expressions for $[X]$ from Equation (4.2) and the expressions for $[A]$ and $[A]^{-1}$ given above. Then use the fact that

$$[X'] = \left\{[A][X][A]^{-1} - [A]\frac{d[A]^{-1}}{dt}\right\} \tag{4.21}$$

to prove the result.

Since the elements of the matrix $[X^{'}]$ are independent of time, the solution of Equation (4.12) is given by

$$\hat{n}^{'}(t) = e^{[Q](t)}\hat{n}^{'}(0), \tag{4.22}$$

where

$$[Q](t) = \int_0^t [X^{'}](t^{'})dt^{'}. \tag{4.23}$$

The matrix $[Q](t)$ is therefore given by

$$[Q](t) = \begin{bmatrix} 0 & (\omega - \omega_0)t & 0 \\ -(\omega - \omega_0)t & 0 & -\omega_1 t \\ 0 & \omega_1 t & 0 \end{bmatrix}. \tag{4.24}$$

Suppose we have the matrix $[S](t)$ which diagonalizes $[Q](t)$, i.e.,

$$[\Lambda](t) = [S]^{-1}(t)[Q](t)[S](t) = diag(\lambda_1(t), \lambda_2(t), \lambda_3(t)), \tag{4.25}$$

where the λ_i (i=1,2,3) are the three eigenvalues of the matrix $[Q](t)$.

Then,

$$\begin{aligned} e^{[Q](t)} &= e^{[S](t)[\Lambda](t)[S]^{-1}(t)} = [I] + [S][\Lambda][S]^{-1} + \frac{1}{2!}([S][\Lambda][S]^{-1})^2 + ..., \\ &= [I] + [S]([\Lambda] + \frac{[\Lambda]^2}{2!} + \frac{[\Lambda]^3}{3!} + ...)[S]^{-1} = [S]e^{[\Lambda](t)}[S]^{-1}. \end{aligned} \tag{4.26}$$

Hence, we get

$$\hat{n}^{'}(t) = [S](t)e^{[\Lambda](t)}[S]^{-1}(t)\hat{n}^{'}(0), \tag{4.27}$$

which leads to

$$\hat{n}(t) = [U]\hat{n}(0) = \left\{[A]^{-1}(t)\left[[S](t)e^{[\Lambda](t)}[S]^{-1}(t)\right]A(0)\right\}\hat{n}(0). \tag{4.28}$$

The last equation describes the time evolution of the spinor on the Bloch sphere in the initial (fixed) frame of reference.

Next, we calculate explicitly the unitary matrix $[U] = [A]^{-1}(t)\left\{[S](t)e^{[\Lambda](t)}[S]^{-1}(t)\right\}[A](0)$. The inverse $[A]^{-1}(t)$ was already calculated above and the matrix [A](0) is simply the 3×3 identity matrix. We are therefore left with calculating the matrix $[S](t)e^{[\Lambda](t)}[S]^{-1}(t)$.

Hereafter, we use the shorthand notations

$$[Q](t) = \begin{bmatrix} 0 & \alpha & 0 \\ -\alpha & 0 & \beta \\ 0 & -\beta & 0 \end{bmatrix}, \tag{4.29}$$

where $\alpha = (\omega - \omega_0)t$ and $\beta = -\omega_1 t$.

The eigenvalues λ of the matrix $[Q](t)$ are found to be

$$\lambda_1 = 0, \tag{4.30}$$

$$\lambda_2 = i\sqrt{\alpha^2 + \beta^2}, \tag{4.31}$$

and

$$\lambda_3 = -i\sqrt{\alpha^2 + \beta^2}. \tag{4.32}$$

For $\lambda_1 = 0$, the corresponding eigenvector is

$$\vec{q_1} = [\beta/\alpha, 0, 1]. \tag{4.33}$$

For $\lambda_2 = i\sqrt{\alpha^2 + \beta^2}$, the corresponding eigenvector is

$$\vec{q_2} = [-\alpha/\beta, -i\sqrt{\alpha^2 + \beta^2}/\beta, 1], \tag{4.34}$$

and for $\lambda_3 = -i\sqrt{\alpha^2 + \beta^2}$, the eigenvector is

$$\vec{q_3} = [-\alpha/\beta, i\sqrt{\alpha^2 + \beta^2}/\beta, 1]. \tag{4.35}$$

Exercises:

- Prove that the eigenvalues of the matrix $Q(t)$ are indeed given by the expressions λ_1, λ_2, and λ_3 given above.
- Derive the expressions of the corresponding eigenvectors, $\vec{q_1}$, $\vec{q_2}$ and $\vec{q_3}$.

These are easy to show and left as exercises for the reader.

It is trivial to show that if we form a matrix $[S]$ whose column vectors are equal to the three eigenvectors of the matrix $[Q]$ found above, then Equation (4.25) will be satisfied. Therefore, the explicit form of the matrix $[S]$ is

$$[S] = \begin{bmatrix} \beta/\alpha & -\alpha/\beta & -\alpha/\beta \\ 0 & -i\sqrt{\alpha^2 + \beta^2}/\beta & i\sqrt{\alpha^2 + \beta^2}/\beta \\ 1 & 1 & 1 \end{bmatrix}, \tag{4.36}$$

and its inverse is given by

$$[S]^{-1} = \begin{bmatrix} \alpha\beta/[\alpha^2 + \beta^2] & 0 & \alpha^2/[\alpha^2 + \beta^2] \\ -0.5\alpha\beta/[\alpha^2 + \beta^2] & 0.5i\beta/[\sqrt{\alpha^2 + \beta^2}] & 0.5\beta^2/[\alpha^2 + \beta^2] \\ -0.5\alpha\beta/[\alpha^2 + \beta^2] & -0.5i\beta/[\sqrt{\alpha^2 + \beta^2}] & 0.5\beta^2/[\alpha^2 + \beta^2] \end{bmatrix}. \tag{4.37}$$

Using the results obtained so far, the matrix $[U] = [A]^{-1}(t)\left\{[S](t)e^{[\Lambda](t)}[S]^{-1}(t)\right\}[A](0)$ appearing on the right hand side of Equation (4.28) is found to be

$$[U] = \begin{bmatrix} g(\delta,\chi)sin\omega t + h(\delta,\chi)cos\omega t & g(\delta,\chi)cos\omega t - cos\delta sin\omega t & [f(\delta)cos\omega t cos\chi - sin\delta sin\omega t]sin\chi \\ -g(\delta,\chi)cos\omega t + h(\delta,\chi)sin\omega t & g(\delta,\chi)sin\omega t + cos\delta cos\omega t & [f(\delta)sin\omega t cos\chi + sin\delta cos\omega t]sin\chi \\ f(\delta)cos\chi sin\chi & -sin\delta sin\chi & cos^2\chi + sin^2\chi cos\delta \end{bmatrix} \tag{4.38}$$

where

$$\begin{aligned}
\delta &= \sqrt{\alpha^2+\beta^2} = \sqrt{(\omega-\omega_0)^2+\omega_1^2 t} \\
f(\delta) &= 1 - cos\delta \\
h(\delta,\chi) &= cos\delta cos^2\chi + sin^2\chi \\
g(\delta,\chi) &= sin\delta cos\chi \\
\chi &= atan[\omega_1/(\omega_0-\omega)] \\
sin^2\chi &= \frac{{\omega_1}^2}{{\omega_1}^2+[\omega_0-\omega]^2}.
\end{aligned} \tag{4.39}$$

4.3 The Rabi formula

Suppose a spinor at time $t = 0$ is at the north pole on the Bloch sphere. We wish to calculate the probability for the spinor to flip to the south pole, as a function of time, when subjected to the combined actions of the static magnetic field B_{old} along the z-axis and the rotating time varying magnetic field B_{new} in the (x-y) plane.

The probability for any arbitrary spinor to be eventually located at the south pole is given by (use Equation (3.26))

$$| < 1|{\xi_n}^+ > |^2 = sin^2(\frac{\theta(t)}{2}) = \frac{1-cos\theta(t)}{2}, \tag{4.40}$$

which is equal to

$$sin^2(\frac{\theta(t)}{2}) = \frac{1-n_z(t)}{2}. \tag{4.41}$$

Now, at $t = 0$, the spinor $\hat{n}(0) = ((n_x(0), n_y(0), n_z(0)) = (0, 0, 1)$. Next we can find $n_z(t)$ from Equation (4.28) and substitute that in the above equation to find

$$sin^2(\frac{\theta(t)}{2}) = \frac{sin^2\chi(t)}{2}[1-cos\delta(t)], \tag{4.42}$$

which is the *Rabi formula*. The quantities χ and δ are defined in Equation (4.39).

The Rabi formula indicates that the probability of spin flipping reaches a maximum of $sin^2\chi$ and therefore can reach unity only if the following resonance condition is satisfied

$$\omega = \omega_0, \tag{4.43}$$

which is mathematically equivalent to the physical discussion at the beginning of this Chapter, i.e., the magnetic field B_{new} in the (x-y) plane must rotate at a frequency equal to the Larmor frequency $g\mu_B B_{old}/\hbar$ in order for the spinor to reach the south pole.

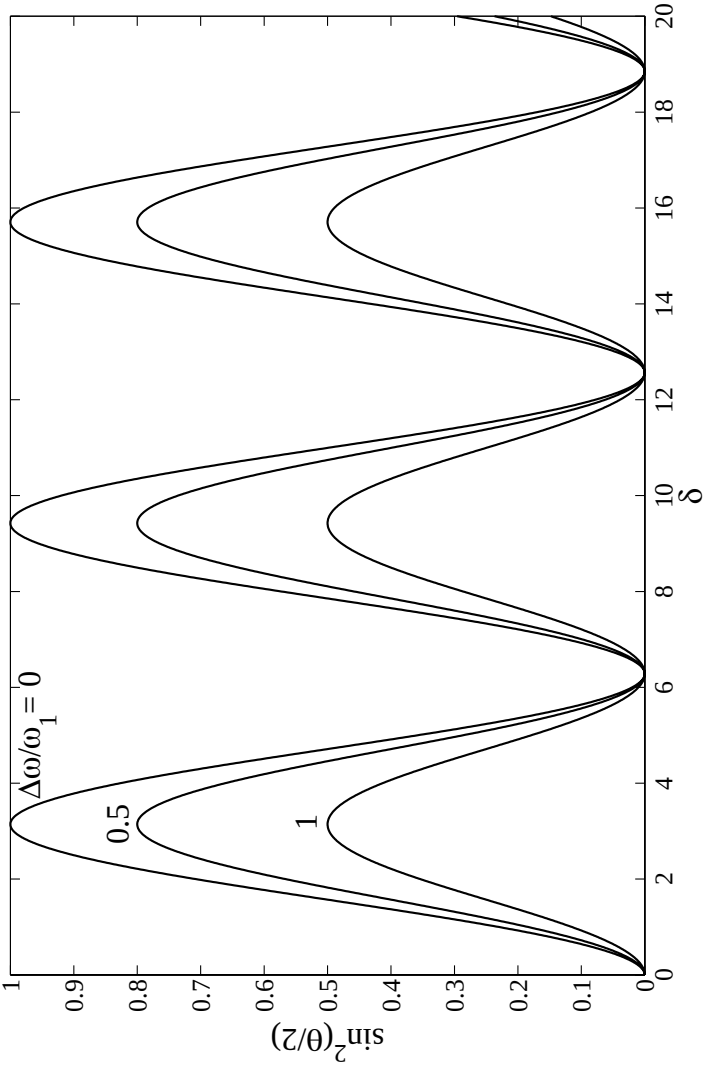


FIGURE 4.3
Plot of the probability for a spinor, initially located at the north pole of the Bloch sphere, to reach the south pole, as a function of δ for various values of the ratio $\Delta\omega/\omega_1$, where $\Delta\omega = \omega - \omega_0$. At resonance, i.e., when $\omega = \omega_0$, the probability to reach the south pole reaches unity periodically. The shortest time to reach the south pole is π/ω_1.

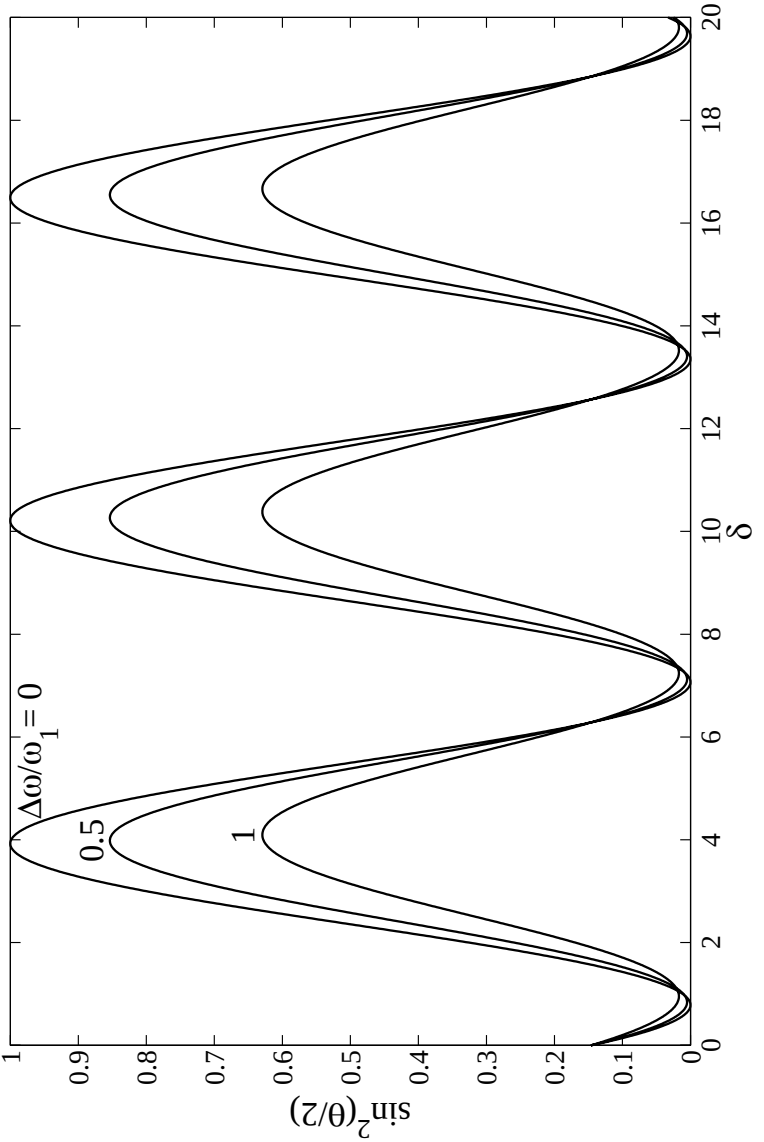


FIGURE 4.4
Time dependence of the probability for a spinor to reach the south pole if initially located in the +x direction on the Bloch sphere. The probability reaches unity at resonance, i.e. when $\omega = \omega_0$, if the phase ϕ of the rotating magnetic field in the (x-y) plane at time $t = 0$ is set equal to ϕ.

4.3.1 Spin flip time

The magnetic field $\vec{B}_{new}$ must rotate in the (x-y) plane with an angular frequency ω_0, which is the Larmor frequency associated with the static field $\vec{B}_{old}$, in order to guarantee a spin flip.

The spinor $\hat{n}$ will periodically visit the north and south poles of the Bloch sphere with a frequency equal to the Larmor frequency of the field in the (x-y) plane, which is ω_1. Therefore, the time that elapses between two consecutive visits of either pole is $T = 2\pi/\omega_1$ and the time required to switch from one pole to the other, which is the time required for the spin to flip, is

$$t_s = \frac{T}{2} = \frac{\pi}{\omega_1} = \frac{\pi\hbar}{g\mu_B B_{new}}. \tag{4.44}$$

This time is inversely proportional to the strength of the field $\vec{B}_{new}$.

The derivation of Rabi's formula was based on the general Equation (4.28) which allows a direct calculation of the time dependence of the probability of a spin transfer between any two points of the Bloch sphere, starting at a location at time t = 0 other than the north pole, which is usually the case derived in most textbooks based on a direct solution of the time-dependent Schrödinger equation.

Example 1: Using the formulation above, calculate the time dependence of the probability (using the same configuration of constant and rotating magnetic fields) for a spinor originally in the eigenstate of the Hadamard matrix with eigenvalue +1 to reach the south pole on the Bloch sphere.

Solution

At t = 0, $((n_x(0), n_y(0), n_z(0)) = (\frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}})$ is the unit vector or Bloch vector associated with the eigenstate of the Hadamard matrix with eigenvalue +1. Using Equation (4.40), we get

$$| < 1|\xi_n{}^+ > |^2 = sin^2(\frac{\theta(t)}{2}) = \frac{\sqrt{2} - 1 + sin\chi(sin\chi - cos\chi)(1 - cos\delta)}{2\sqrt{2}}. \tag{4.45}$$

A plot of that probability as a function of δ is shown in Figure 4.5. Compare it to the case when the initial spinor at time $t = 0$ is at the north pole.

In this case, the probability of reaching the south pole does not ever become unity even when the resonance condition $\omega = \omega_0$ is satisfied. The reason is that the phase of the magnetic field $\vec{B_1}$ in the (x-y) plane must be such that, at time $t = 0$, $\vec{B_1}$ is orthogonal to the plane of the constant magnetic field $\vec{B_0}$ and the original spinor at $t = 0$. Therefore, the phase of the magnetic field $\vec{B_1}$

$$\vec{B_1} = (B_1 cos(\omega t + \phi), B_1 sin(\omega t + \phi), 0) \tag{4.46}$$

must be set to $-\pi/2$ for the spinor to eventually reach the south pole (see Problem 4.7).

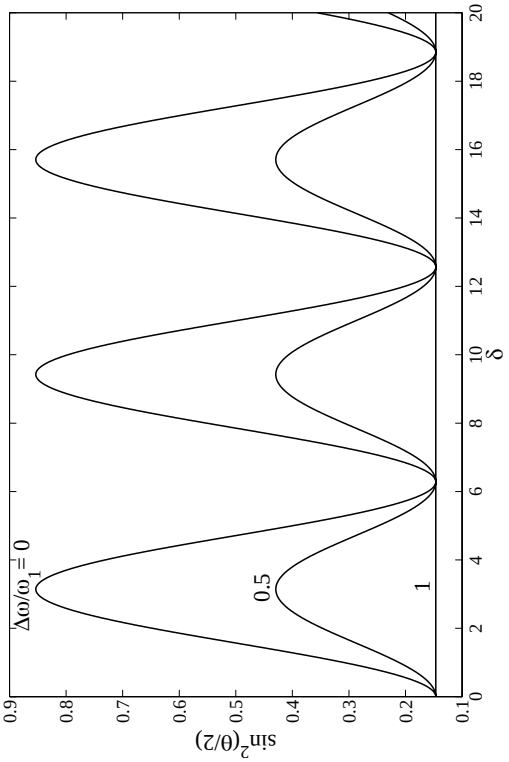


FIGURE 4.5
Time dependence of the probability for a spinor to reach the south pole if initially located on the Bloch sphere at $((n_x(0), n_y(0), n_z(0)) = (\frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}})$, which is an eigenvector of the Hadamard matrix. The probability is maximized, but does ***not*** reach unity, at resonance when $\omega = \omega_0$. In order for the probability to reach unity, the phase ϕ at time t = 0 for the rotating magnetic field in the (x-y) plane must be suitably adjusted (see Problem 4.7).

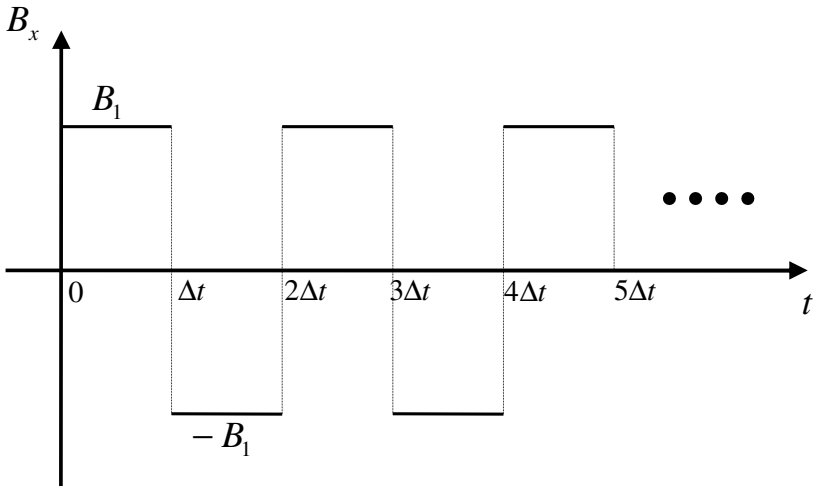


FIGURE 4.6
Waveform of the square wave magnetic field of intensity B_1 applied along the x-axis.

Example 2: Square magnetic pulse along the x-axis

In this section, we investigate the dynamics of spin flipping when the spinor is initially located at the north pole of the Bloch sphere and submitted to the action of a constant magnetic field B_0 along the z-axis, as well as a square wave magnetic field of amplitude B_1 along the x-axis. The waveform of the latter field is shown in Fig. 4.6.

Figures 4.7 through 4.9 show the time dependence of the x, y, and z components of the Bloch vector for the case where the intensities of the magnetic fields are such that $\omega_1/\omega_0 = 0.1$ and ω_1 is positive, i.e., the g-factor for the electron is assumed to be positive. In Figs. 4.7 through 4.9, the time interval Δt is one-half of the period associated with the Larmor frequency ω_0, i.e., the period of the square pulse is exactly the same as the period associated with the Larmor precession due to the magnetic field B_0 alone. In that case, there is a spin flip occurring in a time slightly less than 8 periods. There is also a beat pattern observed in the x and y-components of the Bloch vector corresponding to the qubit spiraling down the surface of the Bloch sphere until it finally reaches the south pole and then spiraling back up towards the north pole. This flipping back and forth between the north and south poles keeps occurring at longer time scales as in the case of a small rotating magnetic field in the (x, y) plane examined above. This case corresponds to the resonant condition when the period of the magnetic field along the x-axis is

the same as the inverse of the Larmor frequency due to B_0 alone.

Figures 4.10 through 4.12 show the time evolution of the the x, y, and z components of the Bloch vector for a qubit initially at the north pole. The period of the magnetic pulse along the x-axis is $1/10^{th}$ of the period associated with the Larmor frequency due to B_0 alone. This case corresponds to the $off - resonance$ case discussed earlier in the description of Rabi oscillations. The magnetic field along the x-axis is changing too frequently, preventing a flip of the spin. Figures 4.10 through 4.12 show that the spin wobbles around the north pole and is never able to leave its proximity.

4.4 Problems

- **Problem 4.1**

  Derive Equation (4.1) starting from the Ehrenfest Theorem.

  Solution

  The Ehrenfest Theorem states that the expectation value (averaged over space) of any time dependent operator $< O >$ obeys the equation

  $$i\hbar\frac{d < O >}{dt} =< [HO - OH] >, \tag{4.47}$$

  where H is the Hamiltonian and the angular brackets denote spatial average.

  For our case, the Hamiltonian is

  $$H = H_0 - (g/2)\mu_B\vec{B}\cdot\vec{\sigma}. \tag{4.48}$$

  Therefore,

  $$i\hbar\frac{d < \sigma_x >}{dt} = -\frac{g\mu_B}{2}\left[B_y(\sigma_y\sigma_x - \sigma_x\sigma_y) + B_z(\sigma_z\sigma_x - \sigma_x\sigma_z)\right]. \tag{4.49}$$

  Next, we use the well known commutation relations of the Pauli spin matrices (see Chapter 2) to obtain

  $$\frac{d < \sigma_x >}{dt} = \frac{g\mu_B}{\hbar}\left(B_y\sigma_z - B_z\sigma_y\right). \tag{4.50}$$

  One can similarly obtain equivalent relations for $d < \sigma_y > /dt$ and $d < \sigma_z > /dt$, to ultimately yield Equation (4.1).

- **Problem 4.2**

  Prove that the inverse of the matrix $[S]$ is indeed given by Equation (4.37).

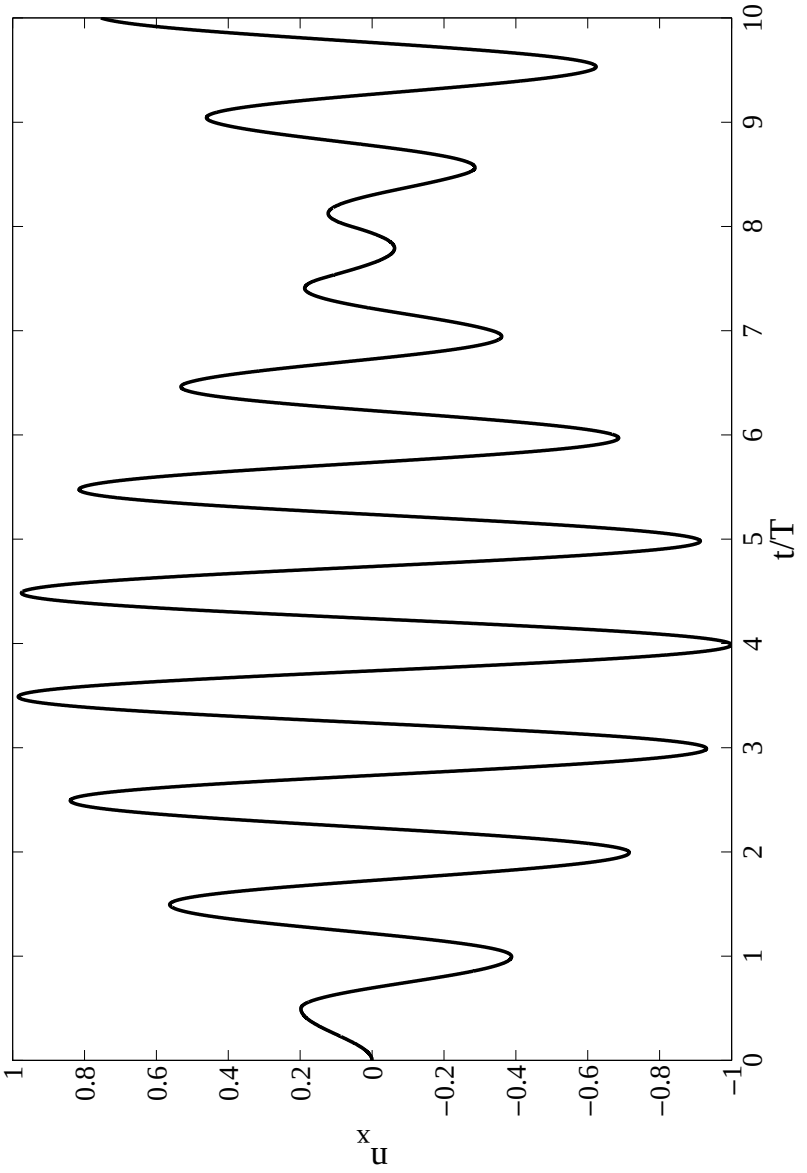


FIGURE 4.7
Time dependence of the x-component of the Bloch vector for a qubit originally at the north pole and submitted to the action of a constant magnetic field along the z-axis and a magnetic field square pulse along the x-axis where the intensities of the two magnetic fields are selected such that $\omega_1/\omega_0 = 0.1$ and ω_1 is positive.

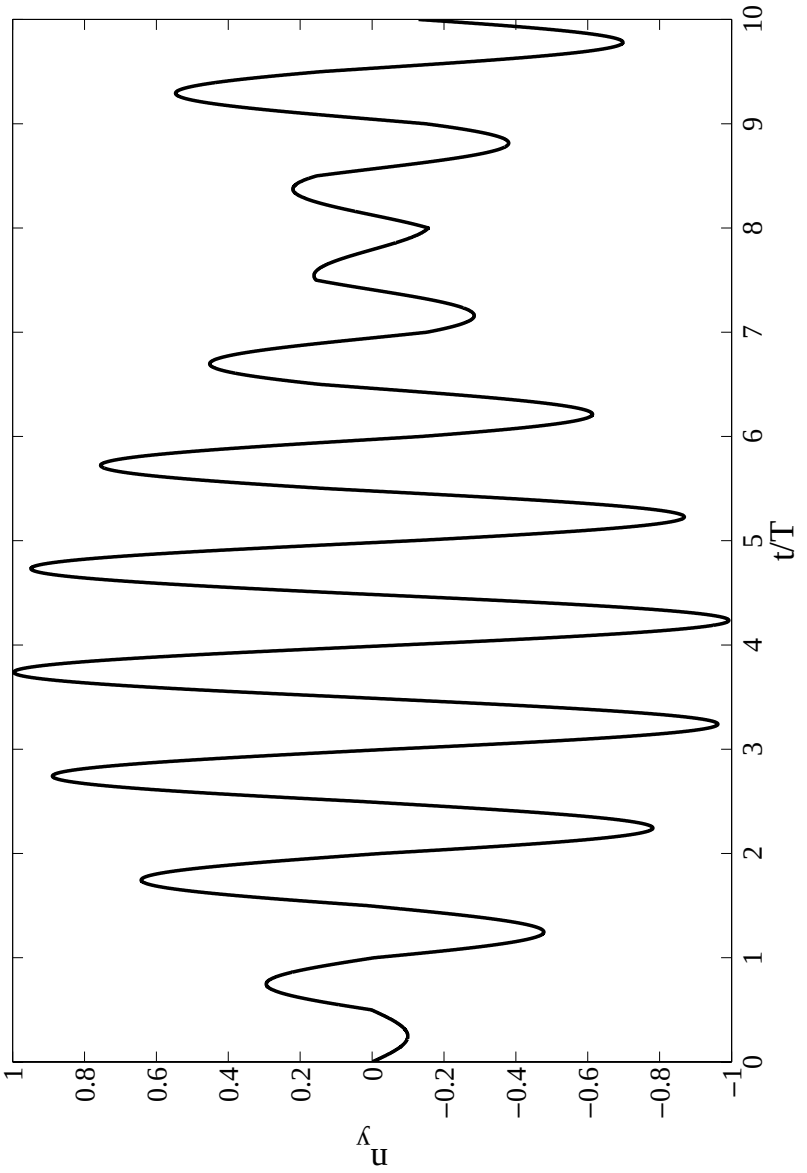


FIGURE 4.8
Same as the previous figure for the y-component of the Bloch vector.

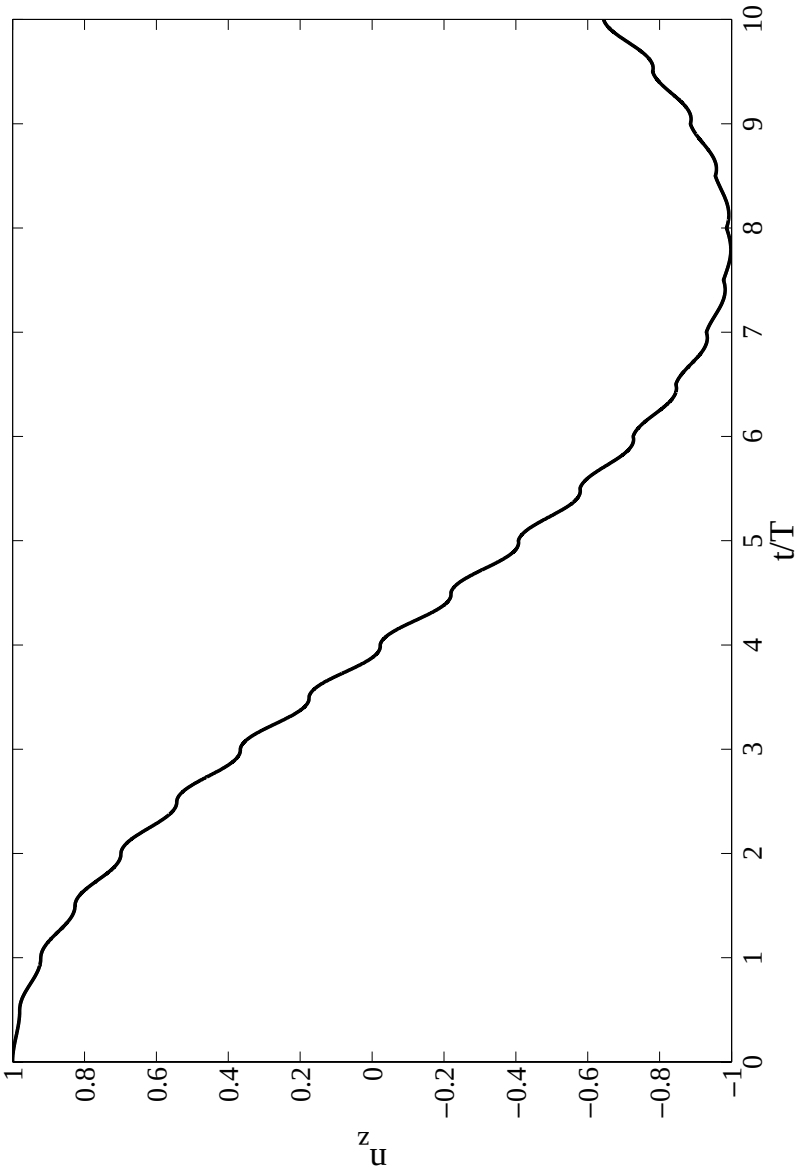


FIGURE 4.9
Same as the previous figure for the z-component of the Bloch vector.

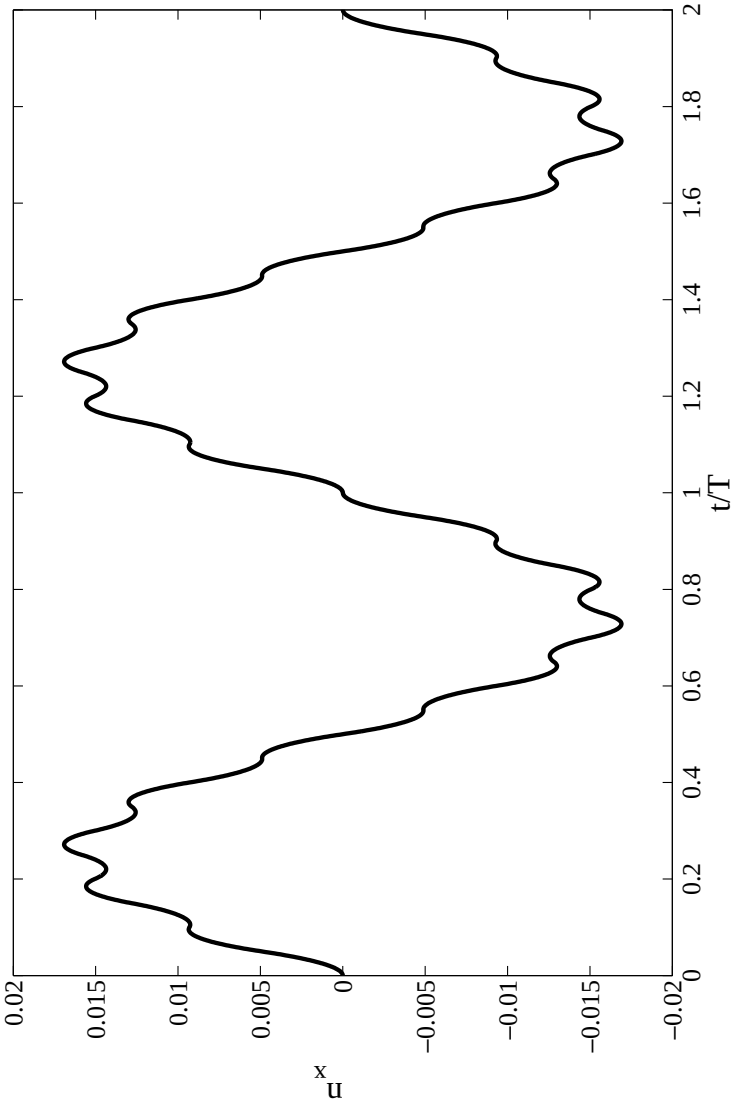


FIGURE 4.10
Time dependence of the x-component of the Bloch vector for a qubit originally at the north pole and submitted to the action of a constant magnetic field along the z-axis and a magnetic field square pulse along the x-axis where the intensities of the two magnetic field are selected such that $\omega_1/\omega_0 = 0.1$ and ω_1 is positive. In this case, the period of the magnetic square pulse along the x-axis is equal to $1/10^{th}$ of the period associated with the Larmor frequency due to the magnetic field B_0 alone.

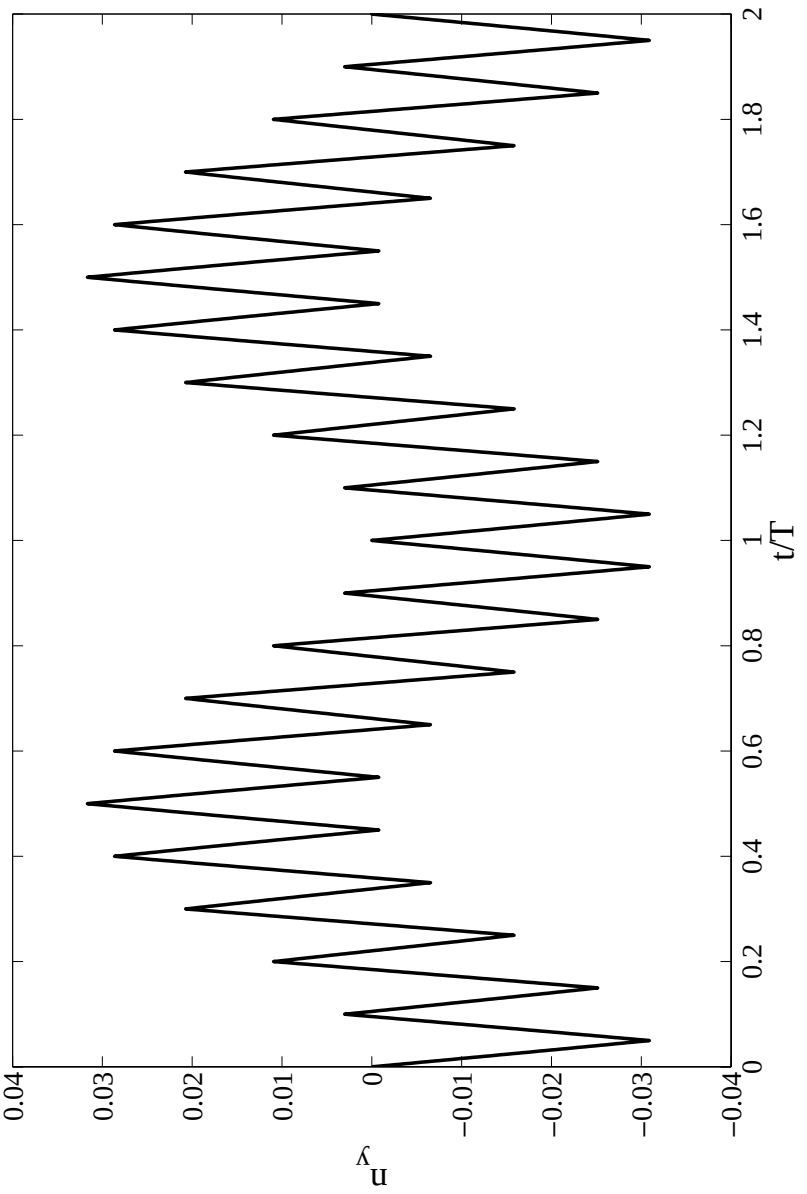


FIGURE 4.11
Same as the previous figure for the y-component of the Bloch vector.

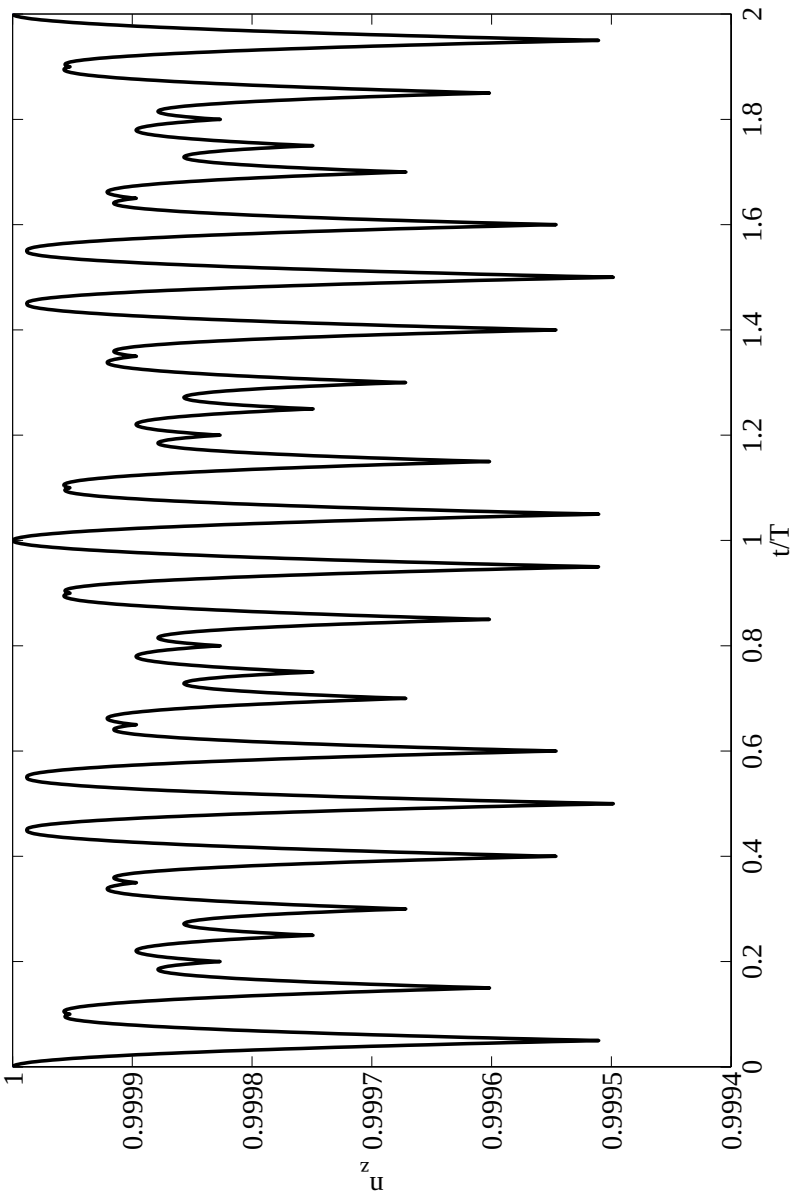


FIGURE 4.12
Same as the previous figure for the z-component of the Bloch vector.

- **Problem 4.3**

  Reproduce the plots shown in Figures (4.3) - (4.5).

- **Problem 4.4**

  Show that the matrix $[U]$ in Equation (4.38) is unitary.

- **Problem 4.5**

  Starting with the general expression for the spinor given in Equation (3.26) in the previous Chapter, show that the probability for the spinor to be aligned along the $+x$-axis at time t is given by

$$| < \xi_{+x} | \xi_n^+ > |^2 = \frac{1 + n_x(t)}{2}. \tag{4.51}$$

  Calculate the explicit time dependence of that probability when the original spinor is located at the north pole at $t = 0$. Repeat for the case when the original spinor is the eigenstate of the Hadamard matrix with eigenvalue $+1$. Find under what condition that probability is maximized.

- **Problem 4.6**

  Starting again with Equation (3.26), show that the probability for the spinor to be aligned along the $+y$-axis at time t is given by

$$| < \xi_{+y} | \xi_n^+ > |^2 = \frac{1 + n_y(t)}{2}. \tag{4.52}$$

  Calculate the explicit time dependence of that probability when the original spinor is located along the $+x$-axis of the Bloch sphere.

- **Problem 4.7**

  In Example 1 discussed above, is there a way to make the probability reach unity by appropriately choosing the intensity of the magnetic field and the phase of the magnetic field rotating in the (x,y) plane? Note that

$$\vec{B} = (B_1 cos(\omega t + \phi), B_1 sin(\omega t + \phi), B_0). \tag{4.53}$$

  Hint: The matrix elements of the unitary matrix in Equation (4.38) are given by

$$\begin{aligned}
u_{11} &= -g_1 sin(\omega t - \varphi) + h_1 cos(\omega t - \varphi) + P, \\
u_{12} &= -g_2 sin(\omega t - \varphi) - h_2 cos(\omega t - \varphi) - Q, \\
u_{13} &= [f(\delta) cos\chi cos(\omega t - \varphi) - sin\delta sin(\omega t - \varphi)]\, sin\chi, \\
u_{21} &= h_1 sin(\omega t - \varphi) + g_1 cos(\omega t - \varphi) + Q,
\end{aligned}$$

$$
\begin{aligned}
u_{22} &= -h_2 sin(\omega t - \varphi) + g_2 cos(\omega t - \varphi) + P, \\
u_{23} &= [f(\delta) cos\chi sin(\omega t - \varphi) + sin\delta cos(\omega t - \varphi)]\, sin\chi, \\
u_{31} &= [f(\delta) cos\chi cos\varphi - sin\delta sin\varphi]\, sin\chi, \\
u_{32} &= -[f(\delta) cos\chi sin\varphi + sin\delta cos\varphi]\, sin\chi, \\
u_{33} &= cos^2\chi + cos\delta sin^2\chi,
\end{aligned} \tag{4.54}
$$

where the following quantities were used:

$$g_1(\delta, \chi, \phi) = sin\varphi cos\delta sin^2\chi, \tag{4.55}$$

$$g_2(\delta, \chi, \phi) = cos\varphi cos\delta sin^2\chi, \tag{4.56}$$

$$h_1(\delta, \chi, \phi) = cos\varphi sin^2\chi, \tag{4.57}$$

$$h_2(\delta, \chi, \phi) = sin\varphi sin^2\chi, \tag{4.58}$$

$$P(\omega t, \delta, \chi) = cos\omega t cos\delta cos^2\chi + sin\omega t sin\delta cos\chi, \tag{4.59}$$

$$Q(\omega t, \delta, \chi) = -cos\omega t sin\delta cos\chi + sin\omega t cos\delta cos^2\chi. \tag{4.60}$$

4.5 References

[1] I. I. Rabi, "On the process of space quantization", Phys. Rev., **49**, 324 (1936).

[2] I. I. Rabi, "Space quantization in a gyrating magnetic field", Phys. Rev., **51**, 652 (1937).

[3] I. I. Rabi, N.F. Ramsey and J. Schwinger, "Use of rotating coordinates in magnetic resonance problems", Rev. Mod. Phys., **26**, 167 (1954).

[4] R. P. Feynman, R.N. Leighton and M. Sands, *The Feynman Lectures of Physics*, Volume II, Chapter IV, Sixth Edition, Addison-Wesley Publishing Company, Massachusetts (1977).

[5] T. F. Havel et al., "Quantum Information Processing by nuclear magnetic resonance spectroscopy", Am. J. Phys., **70**, 345 (2002).

5

The Density Matrix

The density matrix is a useful concept which will allow one to deal with a large collection of spinors with a statistical distribution at time $t = 0$ and follow their time evolution by tracking the time evolution of a new operator (the *density matrix*). This will require the extension of the *Bloch sphere* to the *Bloch ball* concept, in which the vector representing the collection of spinors will be able to reach points in the interior of the Bloch sphere as a function of time. This requires the inclusion of damping and scattering mechanisms representing interactions among spinors, and also their interactions with the environment. The phenomenological equations allowing a description of these relaxation mechanisms are known as **Bloch's equations**. They are introduced here through a connection to the Bloch sphere and Bloch ball concepts. The longitudinal and transverse relaxation times, T_1 and T_2, respectively, are also introduced and some numerical examples (results of Monte-Carlo simulations) are presented to illustrate their physical meaning.

5.1 The density matrix concept: case of a pure state

The concept of density matrix can be introduced starting with the general ket $|\psi\rangle$ describing the quantum state of a system associated with a single particle or with an ensemble of particles, as prescribed in the general axioms of quantum-mechanics (see Chapter 15). To make the concept easier to grasp, we restrict ourselves to a general ket in the two-dimensional complex space $\mathcal{C}_2$, i.e.,

$$|\psi\rangle = \alpha|0> + \beta|1> . \tag{5.1}$$

This state is also referred to as a *pure state* for which the coefficients α and β must satisfy the normalization condition

$$|\alpha|^2 + |\beta|^2 = 1. \tag{5.2}$$

Starting with the ket $|\psi\rangle$, we define a 2×2 *density matrix* as follows

$$\rho = |\psi\rangle\langle\psi| = \begin{pmatrix} \alpha \\ \beta \end{pmatrix} (\alpha^* \beta^*) . \tag{5.3}$$

Hence,

$$\rho = \begin{pmatrix} \alpha\alpha^* & \alpha\beta^* \\ \beta\alpha^* & \beta\beta^* \end{pmatrix} = \begin{pmatrix} |\alpha|^2 & \alpha\beta^* \\ \beta\alpha^* & |\beta|^2 \end{pmatrix}. \tag{5.4}$$

5.2 Properties of the density matrix

The following properties of this matrix are easily derived:

1. It is definite positive, i.e., for any other ket in $\mathcal{C}_2$,

$$< \psi^{'} | \rho | \psi^{'} > \geq 0. \tag{5.5}$$

2. It is Hermitian and, by virtue of property 1, its eigenvalues are therefore real and positive.

3. The trace of the matrix is unity (this results from the normalization of the spinor).

4. Furthermore, $\rho^2 = \rho$.

   Proof:

$$\rho^2 = \begin{pmatrix} \alpha\alpha^* & \alpha\beta^* \\ \beta\alpha^* & \beta\beta^* \end{pmatrix} \begin{pmatrix} \alpha\alpha^* & \alpha\beta^* \\ \beta\alpha^* & \beta\beta^* \end{pmatrix}. \tag{5.6}$$

   Hence

$$\rho^2 = \begin{pmatrix} |\alpha|^4 + |\alpha|^2|\beta|^2 & \alpha\beta^*|\alpha|^2 + \alpha\beta^*|\beta|^2 \\ \beta\alpha^*|\alpha|^2 + \beta\alpha^*|\beta|^2 & |\alpha|^2|\beta|^2 + |\beta|^4 \end{pmatrix}, \tag{5.7}$$

   leading to

$$\rho^2 = \begin{pmatrix} |\alpha|^2 \left(|\alpha|^2 + |\beta|^2\right) & \alpha\beta^* \left(|\alpha|^2 + |\beta|^2\right) \\ \beta\alpha^* \left(|\alpha|^2 + |\beta|^2\right) & |\beta|^2 \left(|\alpha|^2 + |\beta|^2\right) \end{pmatrix}. \tag{5.8}$$

   Taking into account the normalization condition (5.2) for the original spinor, we get

$$\rho^2 = \rho. \tag{5.9}$$

For the case of a spinor described by the ket in $\mathcal{C}_2$ given by Equation (5.1), the average values of the individual spin components are given by

$$< \psi | S_x | \psi > = \frac{\hbar}{2} \left(\alpha\beta^* + \beta\alpha^*\right), \tag{5.10}$$

$$< \psi|S_y|\psi >= \frac{\hbar}{2i}\left(\beta\alpha^* - \alpha\beta^*\right), \tag{5.11}$$

and

$$< \psi|S_z|\psi >= \frac{\hbar}{2}\left(|\alpha|^2 - |\beta|^2\right). \tag{5.12}$$

These results can also be obtained as follows using the definition of the density matrix and the explicit form of the Pauli matrices (where Tr stands for the trace operator)

$$\langle S_x\rangle = \mathtt{Tr}(\rho S_x) = \frac{\hbar}{2}\mathtt{Tr}\left[\begin{pmatrix}\alpha\alpha^* & \alpha\beta^*\\ \beta\alpha^* & \beta\beta^*\end{pmatrix}\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}\right], \tag{5.13}$$

$$\langle S_x\rangle = \frac{\hbar}{2}\left(\alpha\beta^* + \beta\alpha^*\right), \tag{5.14}$$

$$\langle S_y\rangle = \mathtt{Tr}(\rho S_y) = \frac{\hbar}{2}\mathtt{Tr}\left[\begin{pmatrix}\alpha\alpha^* & \alpha\beta^*\\ \beta\alpha^* & \beta\beta^*\end{pmatrix}\begin{pmatrix}0 & -i\\ i & 0\end{pmatrix}\right], \tag{5.15}$$

$$\langle S_y\rangle = \frac{\hbar}{2i}\left(\beta\alpha^* - \alpha\beta^*\right), \tag{5.16}$$

and

$$\langle S_z\rangle = \mathtt{Tr}(\rho S_z) = \frac{\hbar}{2}\mathtt{Tr}\left[\begin{pmatrix}\alpha\alpha^* & \alpha\beta^*\\ \beta\alpha^* & \beta\beta^*\end{pmatrix}\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}\right], \tag{5.17}$$

$$\langle S_z\rangle = \frac{\hbar}{2}\left(|\alpha|^2 - |\beta|^2\right). \tag{5.18}$$

So, the calculation of the average value of any component of the spin operators can be readily obtained from a product of the density matrix and the corresponding Pauli matrix.

Using the above results, we get that for any pure state in $\mathcal{C}_2$

$$< S_x >^2 + < S_y >^2 + < S_z >^2 = \frac{\hbar^2}{4}, \tag{5.19}$$

at all time.

Exercises:

- Prove that the average value $< M > = < \psi|M|\psi >$ of any operator M operating in $\mathcal{C}_2$ can be calculated as follows:

$$< M >= Tr(\rho M). \tag{5.20}$$

The calculation of the average of an operator using Equation (5.20) is actually valid in any Hilbert space.

Solution

In $\mathcal{C}_2$, by definition

$$\begin{aligned}\langle M\rangle &= (\alpha^\star\beta^\star)\begin{bmatrix}M_{11}\ M_{12}\\ M_{21}\ M_{22}\end{bmatrix}\begin{pmatrix}\alpha\\ \beta\end{pmatrix},\\ &= M_{11}|\alpha|^2 + M_{12}\alpha^\star\beta + M_{21}\beta^\star\alpha + M_{22}|\beta|^2. \end{aligned} \tag{5.21}$$

On the other hand

$$\begin{aligned}\mathtt{Tr}(\rho M) &= \mathtt{Tr}\left[\begin{pmatrix}\alpha\alpha^\star\ \alpha\beta^\star\\ \beta\alpha^\star\ \beta\beta^\star\end{pmatrix}\begin{pmatrix}M_{11}\ M_{12}\\ M_{21}\ M_{22}\end{pmatrix}\right],\\ &= M_{11}|\alpha|^2 + M_{12}\alpha^\star\beta + M_{21}\alpha\beta^\star + M_{21}\beta^\star\alpha + M_{22}|\beta|^2. \end{aligned} \tag{5.22}$$

Hence,

$$\langle M\rangle = \langle\psi|M|\psi\rangle = \mathtt{Tr}(\rho M), \tag{5.23}$$

for any 2×2 matrix.

- What is the form of the density matrix ρ associated with the most general form of the spinor on the Bloch sphere which was introduced in Chapter 3? Recall that the spinor is

$$|\psi\rangle = |\xi_n^\dagger\rangle = \cos\frac{\theta}{2}|0\rangle + \sin\frac{\theta}{2}e^{i\phi}|1\rangle. \tag{5.24}$$

By definition:

$$\rho(\theta,\phi) = |\psi\rangle\langle\psi| = \begin{pmatrix}\cos\frac{\theta}{2}\\ \sin\frac{\theta}{2}e^{i\phi}\end{pmatrix}\left(\cos\frac{\theta}{2}, \sin\frac{\theta}{2}e^{-i\phi}\right). \tag{5.25}$$

Hence

$$\rho(\theta,\phi) = \begin{pmatrix}\cos^2\frac{\theta}{2} & \cos\frac{\theta}{2}\sin\frac{\theta}{2}e^{-i\phi}\\ \cos\frac{\theta}{2}\sin\frac{\theta}{2}e^{i\phi} & \sin^2\frac{\theta}{2}\end{pmatrix}. \tag{5.26}$$

- Show explicitly that $\rho(\theta,\phi)$ satisfy the four properties of the density matrix listed in section 2.5.2.

- Is there a state of the form

$$|\psi\rangle = \alpha|+\rangle + \beta|-\rangle, \tag{5.27}$$

which represents a non-polarized spin, i.e., a state for which

$$\langle S_x \rangle = \langle S_y \rangle = \langle S_z \rangle = 0? \tag{5.28}$$

Solution

If we want $\langle S_x \rangle = 0$, we must select α and β such that $\alpha^* \beta$ is purely imaginary.
For $\langle S_y \rangle = 0$, we must have $\alpha^* \beta$ purely real. From these requirements, we must have

$$\alpha^* \beta = 0. \tag{5.29}$$

So, either $\alpha = 0$ and $|\beta| = 1$ and $\langle S_z \rangle = -\frac{\hbar}{2}$, or $\beta = 0$ and $|\alpha| = 1$ and $\langle S_z \rangle = +\frac{\hbar}{2}$.
Hence, $\langle S_z \rangle$ cannot be zero simultaneously with $\langle S_x \rangle$ and $\langle S_y \rangle$. The state $|\psi\rangle = \alpha|+\rangle + \beta|-\rangle$ cannot represent a non-polarized state. This observation will lead us in section 5.3 to the generalization of the concept of density matrix to represent a mixture of quantum states (including a non-polarized ensemble of spinors).

- The time evolution of the density matrix can be obtained using the definition of ρ and using the fact that the Schrödinger equation must be satisfied. This leads to

$$\frac{d\rho}{dt} = \frac{d}{dt}\left[|\psi\rangle\langle\psi|\right] = \left[\left(\frac{d}{dt}|\psi\rangle\right)\langle\psi| + |\psi\rangle\left(\frac{d}{dt}\langle\psi|\right)\right], \tag{5.30}$$

or

$$\frac{d\rho}{dt} = \frac{-i}{\hbar}\left[H|\psi\rangle\langle\psi| - |\psi\rangle\langle\psi|H\right] = \frac{-i}{\hbar}\left[H, \rho\right], \tag{5.31}$$

where the square braket [...] denotes the commutator. The above equation is called the **Liouville equation**.

- Starting with the equation for the time evolution of the density operator above and the calculation of the average of an operator using Equation (5.20), calculate the time evolution of the average value of an operator M and rederive Ehrenfest's equation which is proved in Chapter 15.

Solution

Taking the time derivative of $\mathtt{Tr}(\rho M)$, we get

$$\begin{aligned}\frac{d\langle M\rangle}{dt} &= \frac{d}{dt}\mathtt{Tr}(\rho M), \\ &= \mathtt{Tr}\left(\frac{d\rho}{dt}M\right) + \mathtt{Tr}\left(\rho\frac{dM}{dt}\right).\end{aligned} \tag{5.32}$$

Using Equation (5.31) above, we get

$$\begin{aligned}\frac{d\langle M\rangle}{dt} &= -\frac{i}{\hbar}\mathtt{Tr}\left[[H,\rho]M\right] + \left\langle\frac{dM}{dt}\right\rangle,\\ &= -\frac{i}{\hbar}\mathtt{Tr}\left[H\rho M - \rho H M\right] + \left\langle\frac{dM}{dt}\right\rangle\\ &\quad -\frac{i}{\hbar}\mathtt{Tr}\left[\rho M H - \rho H M\right] + \left\langle\frac{dM}{dt}\right\rangle. \end{aligned} \tag{5.33}$$

Using the properties of the trace, $\mathtt{Tr}(AB) = \mathtt{Tr}(BA)$,

$$\frac{d<M>}{dt} = -\frac{i}{\hbar}\mathtt{Tr}\left[\rho[M,H]\right] + \left\langle\frac{dM}{dt}\right\rangle, \tag{5.34}$$

or

$$\frac{d<M>}{dt} = -\frac{i}{\hbar}\left\langle[M,H]\right\rangle + \left\langle\frac{dM}{dt}\right\rangle, \tag{5.35}$$

which is Ehrenfest's theorem proved in a different way in Chapter 15.

5.3 Pure versus mixed state

Example 1:

Starting with Equation (5.26) for the density matrix associated with a general spinor on the Bloch sphere and performing an average over the angles assuming a uniform distribution of the spinors on the Bloch sphere, we generate a new density matrix:

$$\rho = \frac{1}{4\pi}\int d\Omega\rho(\theta,\varphi) = \frac{1}{4\pi}\int_0^{2\pi} d\varphi\int_0^{\pi}\sin\theta d\theta\rho(\theta,\varphi). \tag{5.36}$$

Performing the integration, we get $\rho = \begin{pmatrix}\frac{1}{2} & 0\\ 0 & \frac{1}{2}\end{pmatrix}$. It is easy to show that the four properties listed in section 5.2 are satisfied, except for $\rho^2 = \rho$. In fact

$$\rho^2 = \begin{pmatrix}\frac{1}{4} & 0\\ 0 & \frac{1}{4}\end{pmatrix}, \tag{5.37}$$

and Tr $(\rho^2) = 0.5$.

Furthermore, if we compute the average values of the spin components using the Equations (5.13), (5.15) and (5.18), we find that the three average values are $\langle S_i\rangle = \mathtt{Tr}(\rho S_i) = 0$, for $i = x, y, z$.

So, the density matrix obtained by averaging over the angles seems to represent a collection of spinors uniformly distributed on the Bloch sphere. For

that distribution, we do not expect an average magnetic moment for the collection of spinors. Indeed, we found that the ensemble averages of the three components of the spin operators were each equal to zero.

Example 2:

- Starting with the following normalized kets

$$|a_1>=\sqrt{\frac{3}{5}}|0>+\sqrt{\frac{2}{5}}|1>, \tag{5.38}$$

and

$$|a_2>=\sqrt{\frac{3}{5}}|0>-\sqrt{\frac{2}{5}}|1> \tag{5.39}$$

(which are not orthogonal), consider the operator ρ defined as follows

$$\rho=\frac{1}{2}|a_1><a_1|+\frac{1}{2}|a_2><a_2|. \tag{5.40}$$

Show that

$$\rho=\frac{3}{5}|0><0|+\frac{2}{5}|1><1|. \tag{5.41}$$

Then, show that

$$Tr(\rho)=1. \tag{5.42}$$

Calculate $Tr(\rho^2)$ and show that it is less than 1.

- Show that $<S_x>=<S_y>=0$, and $<S_z>=\frac{\hbar}{4}$ and therefore, $<S_x>^2+<S_y>^2+<S_z>^2=\frac{\hbar^2}{16}$ which is less than the value for the case of a pure state.

Equation (5.41) looks like a superposition of the density matrices associated with the spinors $|0>$ and $|1>$ with the additional property that the sum of the coefficients in front of the individual matrices is unity.

We can generalize the results of the previous example by considering a density matrix defined using a set of normalized spinors on the Bloch sphere, i.e.,

$$\rho=\sum_j p_j|\psi_j\rangle\langle\psi_j|=\sum_j p_j\rho_j, \tag{5.43}$$

where the $p_i's$ are the probabilities to find a spinor in state $|\psi_j\rangle$ and ρ_j is the density matrix associated with state $|\psi_j\rangle$. If the sum contains more than two terms, the $|\psi_j\rangle$'s cannot be orthogonal to one another. Furthermore, since the $p_j's$ are probabilities, the sum over the index j of these coefficients appearing in the general expression of the density matrix associated with a collection of spinors must be unity.

- For the density matrix ρ defined by Equation (5.43), show that the properties 1 through 3 in section 5.2 are still satisfied.

- Using the general expression (5.43) for ρ, show that Tr $(\rho^2) < 1$.

Let us now compute $\langle \overrightarrow{V} \rangle = Tr[\rho \overrightarrow{V}]$, which is the generalized Bloch vector associated with a collection of spinors described by a density matrix of the form (5.43).

Solution

$$\begin{aligned}
\langle \overrightarrow{V} \rangle = \mathtt{Tr}[\rho \overrightarrow{V}] &= \mathtt{Tr}\left[\sum_j p_j |\psi_j\rangle\langle\psi_j| \overrightarrow{V}\right], \\
&= \sum_j p_j \mathtt{Tr}\left[|\psi_j\rangle\langle\psi_j| \overrightarrow{V}\right], \\
&= \sum_j p_j Tr[\langle\psi_j| \overrightarrow{V} |\psi_j\rangle], \\
&= \sum_j p_j \langle\psi_j| \overrightarrow{V} |\psi_j\rangle, \\
&= \sum_j p_j \overrightarrow{v_j}, \qquad (5.44)
\end{aligned}$$

where $\overrightarrow{v_j} = \langle\psi_j| \overrightarrow{V} |\psi_j\rangle$ and the summation is a vector sum. Hence, the generalized Bloch vector $\langle \overrightarrow{V} \rangle$ corresponding to an ensemble of pure states $|\psi_j\rangle$ is the weighted vector sum of the Bloch vectors representing the members of the ensemble.

From a physical point of view, the quantity $\langle \overrightarrow{V} \rangle$ is intuitively appealing since the weighted sum $\sum_j p_j \overrightarrow{v_j}$ is proportional to the average magnetic moment associated with the collection of spinors. The next question to address is to find out the location of the vector $\langle \overrightarrow{V} \rangle$ with respect to the Bloch sphere.

Exercises:

- Show that for any equally-weighted two-member ensemble with $\langle\psi_j|\psi_j\rangle = 0$, the ensemble Bloch vector will be zero if $|\psi_1\rangle$ and $|\psi_2\rangle$ correspond to antipodal points on the Bloch sphere.

  Solution

  For such an ensemble, we have $\rho = \frac{1}{2}\left(|\psi_1\rangle\langle\psi_1| + |\psi_2\rangle\langle\psi_2|\right) = \frac{1}{2}\mathbf{1}$, and therefore $\langle \overrightarrow{V} \rangle = Tr(\rho \overrightarrow{V}) = 0$.

- Consider two ensembles of spinors, one for which the spinors are in the states below with the corresponding probabilities

$$\begin{aligned}
&|\psi_1\rangle \ , \ p_1 = 0.5, \\
&|\psi_2\rangle \ , \ p_2 = 0.5. \qquad (5.45)
\end{aligned}$$

The second ensemble is composed of spinors in the following states with the associated probabilities

$$|\psi_+\rangle = \frac{|\psi_1\rangle + |\psi_1\rangle}{\sqrt{2}}\ , p_1 = 0.5,$$
$$|\psi_-\rangle = \frac{|\psi_1\rangle - |\psi_1\rangle}{\sqrt{2}}\ , p_2 = 0.5. \tag{5.46}$$

Show that the density matrices associated with the two mixed states described above are identical. Since the density matrix completely characterizes the state of a system, this last exercise shows that two seemingly different mixtures of states can be physically equivalent. In other words, a given density matrix can describe a large number of different mixed states.

- Which of the following matrices qualify as density matrices? Why?

$$\rho_1 = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}, \tag{5.47}$$

$$\rho_2 = \begin{bmatrix} .25 & 1 \\ 1 & .75 \end{bmatrix}, \tag{5.48}$$

$$\rho_3 = \begin{bmatrix} .25 & 0.2 + i0.1 \\ 0.2 - i0.1 & .75 \end{bmatrix}, \tag{5.49}$$

and

$$\rho_4 = \begin{bmatrix} 0.5 & .25 \\ .25 & 0.5 \end{bmatrix}. \tag{5.50}$$

For those which do not qualify, state the reason(s) why.

- In the previous problem, which of the density matrices are associated with a pure state? a mixed state? State the reason why.

5.4 Concept of the Bloch ball

In the previous sections and exercises, the density matrices associated with pure or mixed states were found to have the general form

$$\rho = \begin{bmatrix} a + b & c - id \\ c + id & a - b \end{bmatrix}, \tag{5.51}$$

where (a, b, c, d) are real.

This form is just another way to restate the theorem proven in Chapter 2 that any matrix M associated with a Hermitian operator can be decomposed in terms of the 2×2 identity and spin Pauli matrices. Indeed, since $Tr(\rho) = 1$ for density matrices associated with pure and mixed states, we must have $a = \frac{1}{2}$ and Equation (5.51) can be rewritten as:

$$\rho = \frac{1}{2}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + b\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} + c\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} + d\begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \tag{5.52}$$

or, in a more compact form

$$\rho = \frac{1}{2}[I + 2c\sigma_x + 2d\sigma_y + 2b\sigma_z]. \tag{5.53}$$

Using this last expression, the coefficients (b, c, d) can be easily related to the average components of the spin operators. In fact,

$$< \sigma_x >= Tr(\rho\sigma_x) = 2c, < \sigma_y >= Tr(\rho\sigma_y) = 2d, < \sigma_z >= Tr(\rho\sigma_z) = 2b. \tag{5.54}$$

Hence, we have

$$\rho = \frac{1}{2}[I + \overrightarrow{\nu} \cdot \vec{\sigma}], \tag{5.55}$$

where $\vec{\nu} = (\nu_x, \nu_y, \nu_z) = (< \sigma_x >, < \sigma_y >, < \sigma_z >)$.

A density matrix corresponds to a pure state when $Tr(\rho)^2 = 1$, i.e., when

$$Tr(\rho)^2 = \frac{1}{4}[(1 + \nu_z)^2 + (1 - \nu_z)^2 + 2{\nu_x}^2 + 2{\nu_y}^2] = \frac{1}{2}[1 + \vec{\nu} \cdot \vec{\nu}] = 1, \tag{5.56}$$

or whenever $|\vec{\nu}| = 1$.

Since ρ is a Hermitian operator which must be positive, all its eigenvalues are positive. Solving for the eigenvalues of the general density matrix (5.51), we find

$$\lambda = \frac{1}{2} \pm \sqrt{b^2 + c^2 + d^2}. \tag{5.57}$$

For $\lambda \geq 0$, we must have $\sqrt{b^2 + c^2 + d^2} \leq 1/2$. This implies that $|\vec{\nu}| \leq 1$.

This last property shows that unless the norm $|\vec{\nu}|$ is equal to unity, the generalized Bloch vector $\vec{\nu}$ associated with a mixed state will represent a point inside the Bloch sphere. In other words, these points can only be reached as representative of a state of a collection of particles whose density matrix correspond to a mixed state. The ensemble of these points which are within the Bloch sphere makes up the *Bloch ball*.

Exercises

- Consider the matrix A

$$A = \alpha|0 >< 0| + \beta|0 >< 1| + \delta|1 >< 0| + \gamma|1 >< 1|, \tag{5.58}$$

where α, β, γ and δ are complex numbers. What are the conditions on these parameters for the matrix A to be the density matrix associated with a pure state? mixed state?

- For each of the density matrices listed in the third exercise at the end of section 5.3, calculate the component of the vector $\vec{\nu}$ characterizing the vector representation of the pure or mixed state on the Bloch sphere or inside the Bloch ball.

- Consider the Bloch ball in Fig. 5.1. For the points A, B, C, D, E marked on or inside the Bloch ball, what is the density matrix operator associated with that point? The Cartesian coordinates of the points are A: (1.0,0.0,0.0), B: (0.0,1.0,0.0), C: (0.0,0.0,0.5), D: $(1/\sqrt{3},1/\sqrt{3},1/\sqrt{3})$, and E: $(1/\sqrt{2},1/\sqrt{2},0.0)$.

5.5 Time evolution of the density matrix: case of mixed state

Assuming we have an ensemble of spinors represented by a density operator ρ given by Equation (5.43), we can derive its equation of motion in the following way:

$$\begin{aligned}
\frac{d\rho}{dt} &= \frac{d}{dt}\left[\sum_j p_j|\psi_j\rangle\langle\psi_j|\right], \\
&= \sum_j p_j\left[\left(\frac{d}{dt}|\psi_j\rangle\right)\langle\psi_j| + |\psi_j\rangle\left(\frac{d}{dt}|\psi_j\rangle\right)\right], \\
&= \sum_j p_j\left[\left(\frac{-i}{\hbar}H|\psi_j\rangle\right)\langle\psi_j| + |\psi_j\rangle\left(\frac{i}{\hbar}H|\psi_j\rangle\right)\right], \\
&= \frac{-i}{\hbar}\sum_j p_j\left[H|\psi_j\rangle\langle\psi_j| - |\psi_j\rangle\langle\psi_j|H\right], \\
&= \frac{-i}{\hbar}\left[H,\rho\right], \qquad (5.59)
\end{aligned}$$

which is the **Liouville Equation**.

In this derivation, the Hamiltonian H is the form of the Hamiltonian associated with a single spinor and the time evolution of the density matrix for a mixed state is found to be equivalent to that for a pure state.

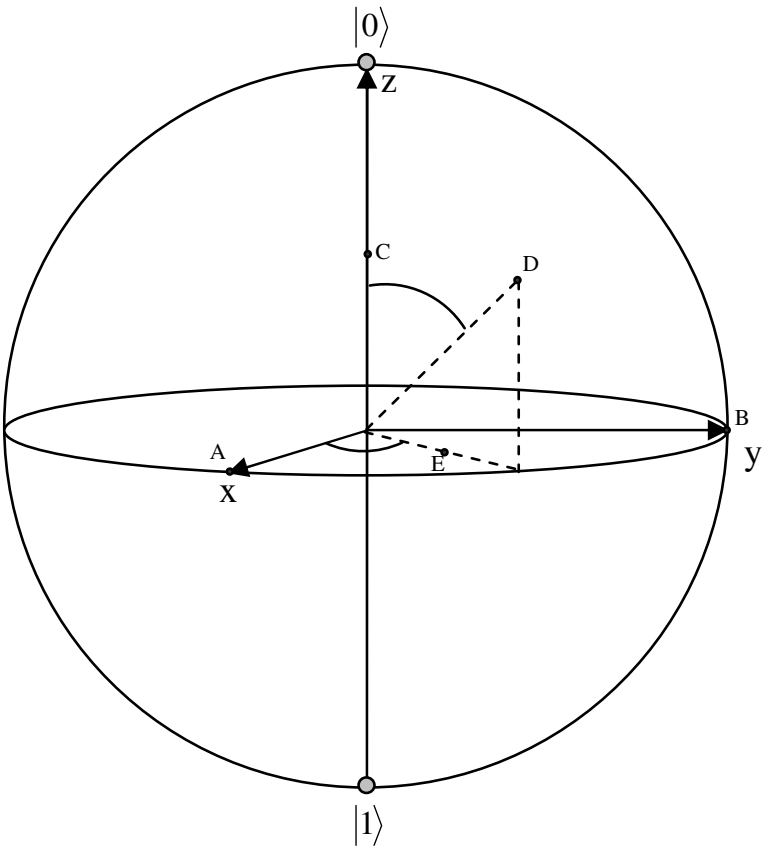


FIGURE 5.1
The Bloch ball.

Similarly, we know that, from postulate 2 of quantum mechanics discussed in Chapter 15,

$$|\psi_j(t)\rangle = U(t)|\psi_j(0)\rangle, \tag{5.60}$$

where $U(t)$ is a unitary operator and therefore

$$\langle\psi_j(t)| = \langle\psi_j(0)|U^\dagger(t). \tag{5.61}$$

Therefore

$$\begin{aligned}\rho(t) &= \sum_j p_j U(t)|\psi_j(0)\rangle\langle\psi_j(0)|U^\dagger(t),\\ &= U(t)\left[\sum_j p_j|\psi_j(0)\rangle\langle\psi_j(0)|\right]U^\dagger(t),\\ &= U(t)\rho(0)U^\dagger(t).\end{aligned} \tag{5.62}$$

Hence

$$Tr\left[\rho^2(t)\right] = Tr\left[U(t)\rho^2(0)U^\dagger(t)\right] = Tr\left[U(t)\rho(0)\rho(0)U^\dagger(t)\right]. \tag{5.63}$$

Using the properties of the trace, we get

$$Tr\left[\rho^2(t)\right] = Tr\left[\rho(0)U^\dagger(t)U(t)\rho(0)\right] = Tr\left[\rho(0)\rho(0)\right], \tag{5.64}$$

and therefore,

$$Tr\left[\rho^2(t)\right] = Tr\left[\rho^2(0)\right]. \tag{5.65}$$

So, for a system characterized by a Hamiltonian (Hermitian matrix), the time evolution preserves the length of the generalized Bloch vector. As time evolves, the latter will move inside the Bloch ball but will stay at a fixed distance from its center.

We can use the above results to derive the equations of motion for the generalized Bloch vector components. Using Equation (5.55):

$$\begin{aligned}\frac{d\nu_j}{dt} &= \frac{d}{dt}\left(\mathtt{Tr}[\rho\sigma_j]\right),\\ &= \mathtt{Tr}\left[\sigma_j\frac{d}{dt}\rho\right],\\ &= \frac{-i}{\hbar}\mathtt{Tr}\left[\sigma_j[H,\rho]\right],\\ &= \frac{-i}{2\hbar}\mathtt{Tr}\left[\sigma_j[H,(I+\overrightarrow{\nu}\cdot\overrightarrow{\sigma}]\right].\end{aligned} \tag{5.66}$$

Hence, the evolutions of the three components will generally be coupled.

Example: We seek the explicit form of the differential equations governing the time evolution of the components v_j associated with a spin-1/2 particle with gyromagnetic factor γ in a spatially independent magnetic field $\vec{B}$. In this case, the Hamiltonian for an individual spinor is given by

$$H = -\gamma \vec{S} \cdot \vec{B} = -\gamma \frac{\hbar}{2} (\vec{\sigma} \cdot \vec{B}). \tag{5.67}$$

Equation (5.66) becomes

$$\frac{d\nu_j}{dt} = \frac{-i}{2\hbar} Tr \left[< \sigma_j > \left[-\gamma \frac{\hbar}{2} \vec{\sigma} \cdot \vec{B}, (1 + \vec{\nu} \cdot \vec{\sigma}) \right] \right], \tag{5.68}$$

where, by definition

$$\vec{\sigma} \cdot \vec{B} = \sigma_x B_x + \sigma_y B_y + \sigma_z B_z, \tag{5.69}$$

and

$$\vec{\nu} \cdot \vec{\sigma} = \nu_x \sigma_x + \nu_y \sigma_y + \nu_z \sigma_z. \tag{5.70}$$

Equation (5.68) leads to

$$\frac{d\nu_j}{dt} = \frac{i\gamma}{4} Tr \left[\sigma_j (\vec{\sigma} \cdot \vec{B})(1 + \vec{\nu} \cdot \vec{\sigma}) - \sigma_j (1 + \vec{\nu} \cdot \vec{\sigma})(\vec{\sigma} \cdot \vec{B}) \right]. \tag{5.71}$$

Evaluating the products on the right hand side and regrouping terms, we find

$$\begin{aligned} \frac{d\nu_j}{dt} = \frac{i\gamma}{4} Tr[&\sigma_j \sigma_x \sigma_y (B_x \nu_y - \nu_x B_y) + \sigma_j \sigma_x \sigma_z (B_x \nu_z - \nu_x B_z) \\ &+ \sigma_j \sigma_y \sigma_x (B_y \nu_x - \nu_y B_x)] \\ + \frac{i\gamma}{4} Tr[&\sigma_j \sigma_y \sigma_z (B_y \nu_z - \nu_y B_z) + \sigma_j \sigma_z \sigma_x (B_z \nu_x - \nu_z B_x) \\ &+ \sigma_j \sigma_z \sigma_y (B_z \nu_y - \nu_z B_y)]. \end{aligned} \tag{5.72}$$

Focusing on the x-component first,

$$\begin{aligned} \frac{d\nu_x}{dt} = \frac{i\gamma}{4} Tr[&\sigma_x \sigma_x \sigma_y (B_x \nu_y - \nu_x B_y) + \sigma_x \sigma_x \sigma_z (B_x \nu_z - \nu_x B_z) \\ &+ \sigma_x \sigma_y \sigma_x (B_y \nu_x - \nu_y B_x)] \\ + \frac{i\gamma}{4} Tr[&\sigma_x \sigma_y \sigma_z (B_y \nu_z - \nu_y B_z) + \sigma_x \sigma_z \sigma_x (B_z \nu_x - \nu_z B_x) \\ &+ \sigma_x \sigma_z \sigma_y (B_z \nu_y - \nu_z B_y)], \end{aligned} \tag{5.73}$$

and using the properties of the trace operator

$$\begin{aligned} \frac{d\nu_x}{dt} = \frac{i\gamma}{4} Tr[&\sigma_x \sigma_x \sigma_y (B_x \nu_y - \nu_x B_y) + \sigma_x \sigma_x \sigma_z (B_x \nu_z - \nu_x B_z) \\ &+ \sigma_x \sigma_x \sigma_y (B_y \nu_x - \nu_y B_x)] \\ + \frac{i\gamma}{4} Tr[&\sigma_x \sigma_y \sigma_z (B_y \nu_z - \nu_y B_z) + \sigma_x \sigma_x \sigma_z (B_z \nu_x - \nu_z B_x) \\ &+ \sigma_x \sigma_z \sigma_y (B_z \nu_y - \nu_z B_y)]. \end{aligned} \tag{5.74}$$

Since ${\sigma_x}^2 = 1$, we get

$$\begin{aligned}\frac{d\nu_x}{dt} = &\frac{i\gamma}{4}Tr\left[\sigma_y(B_x\nu_y - \nu_x B_y) + \sigma_z(B_x\nu_z - \nu_x B_z) + \sigma_y(B_y\nu_x - \nu_y B_x)\right] \\ &+\frac{i\gamma}{4}Tr[\sigma_x\sigma_y\sigma_z(B_y\nu_z - \nu_y B_z) \\ &+\sigma_z(B_z\nu_x - \nu_z B_x) + \sigma_x\sigma_z\sigma_y(B_z\nu_y - \nu_z B_y)]. \qquad (5.75)\end{aligned}$$

In this last equation, the first trace is identically zero because $Tr\sigma_x = Tr\sigma_y = Tr\sigma_z = 0$. Hence,

$$\frac{d\nu_x}{dt} = \frac{i\gamma}{4}Tr\left[\sigma_x\sigma_y\sigma_z(B_y\nu_z - \nu_y B_z) + \sigma_x\sigma_z\sigma_y(B_z\nu_y - \nu_z B_y)\right]. \quad (5.76)$$

This leads to

$$\frac{d\nu_x}{dt} = \frac{i\gamma}{4}(B_y\nu_z - \nu_y B_z)Tr\left[\sigma_x\sigma_y\sigma_z - \sigma_x\sigma_z\sigma_y\right], \qquad (5.77)$$

and, since $\sigma_x\sigma_y\sigma_z = iI$ and $\sigma_x\sigma_z\sigma_y = -iI$, we finally get

$$\frac{d\nu_x}{dt} = -\gamma(B_y\nu_z - \nu_y B_z). \qquad (5.78)$$

Proceeding the same way for the other two components ν_y and ν_z, we get

$$\begin{aligned}\frac{d\nu_y}{dt} &= \frac{i\gamma}{4}(B_x v_z - v_x B_z)Tr\left[\sigma_y\sigma_x\sigma_z - \sigma_y\sigma_z\sigma_x\right] \\ &= -\gamma(B_z v_x - v_z B_x), \qquad (5.79)\end{aligned}$$

and

$$\begin{aligned}\frac{d\nu_z}{dt} &= \frac{i\gamma}{4}(B_x v_y - v_x B_y)Tr\left[\sigma_z\sigma_x\sigma_y - \sigma_z\sigma_y\sigma_x\right] \\ &= -\gamma(B_x v_y - v_x B_y). \qquad (5.80)\end{aligned}$$

In vector notation, the last three equations can be written in the very compact form

$$\frac{d\vec{\nu}}{dt} = \gamma\,\overrightarrow{v} \times \overrightarrow{B}. \qquad (5.81)$$

5.6 The relaxation times T_1 and T_2 and the Bloch equations

If we consider a large ensemble of spinors located on the Bloch sphere at time $t = 0$ and submitted to the action of a constant magnetic field B_0 along the

z-axis, the solution of the three equations (5.78, 5.79, 5.80) can easily be found since the differential equations reduce to

$$\frac{d\nu_x}{dt} = -\gamma(B_y\nu_z - \nu_y B_z) = \gamma B_0 \nu_y, \tag{5.82}$$

$$\frac{d\nu_y}{dt} = -\gamma(B_z v_x - v_z B_x) = -\gamma B_0 \nu_y, \tag{5.83}$$

and

$$\frac{d\nu_z}{dt} = -\gamma(B_x v_y - v_x B_y) = 0. \tag{5.84}$$

From these equations, it is easy to show that

$$\nu_x \frac{d\nu_x}{dt} + \nu_y \frac{d\nu_y}{dt} = 0 \tag{5.85}$$

and

$$\frac{d\nu_z}{dt} = 0. \tag{5.86}$$

This means that the generalized Bloch vector representing the collection of spinors rotates on a cone centered around the z-axis with the Larmor frequency. This is expected since each individual spinor rotates on a cone with the Larmor frequency when subjected to the action of a constant magnetic field. Since the relative locations of the spinors stay unchanged, the generalized Bloch vector, which represents a weighted average of their locations, also rotates with the Larmor frequency around the z-axis and its magnitude stays unchanged. For an observer in a frame rotating with the Larmor frequency with respect to a fixed frame of reference of the enclosure containing the spinors, all spinors will appear frozen in time.

In a realistic situation, there will be some relaxation mechanisms which will have two major effects on the evolution of the ensemble of the spinors on the Bloch sphere. Imagine a gas of atoms in a finite enclosure in the presence of a constant magnetic field along the z-axis. Suppose all spinors are at the North pole at time $t = 0$, which corresponds to a high energy state for the Hamiltonian (5.67) when γ is negative. Eventually, some of the spinors will relax towards the south pole by losing energy via interaction with the environment until a distribution of spinors is reached agreeing with a thermodynamic description of the spinors. The population of the low and high energy Zeeman states must be in agreement with a Boltzmann distribution of the spinors characterized by a Bloch vector located inside the Bloch ball:

$$(\nu_x, \nu_y, \nu_z) = (0, 0, \nu_{z0}). \tag{5.87}$$

For the case of an ensemble of spin-1/2 particles in a constant external magnetic field, ν_{z0} is given by the well-known Langevin equation [3].

The timescale for relaxation from the initial distribution of spinors towards the ground state in agreement with the laws of thermodynamics is called the

longitudinal relaxation time T_1 and is a measure of the time required for spins in the higher-energy eigenstate to decay back down to the ground state. The time T_1 is usually associated with energy dissipation. Hence, a long T_1 corresponds to a slow rate of energy loss due to interaction with the environment. For the case of gas atoms mentioned above, one possible mechanism contributing to T_1 is the interaction between individual spins as a result of collisions. As the spin of any given atom in the ensemble is randomly perturbed through collisions with other atoms (which carry their own magnetic moment), this atom will feel a time-dependent magnetic field, whose magnitude and direction will essentially be random owing to the randomness of the atomic spatial trajectories during collisions.

Whatever the time duration of the collisions is, the x and y components of the field pulse will have many Fourier components, one of which will be close in frequency to the Larmor frequency associated with the externally applied magnetic field. This component will trigger collision-induced spin flips, as was discussed in the description of Rabi oscillations in Chapter 4. If the collision time is short compared to the inverse of the Larmor frequency, many collisions will be necessary before a spin flip occurs, and therefore the time T_1 can be relatively large. Also, as was shown in the derivation of the Rabi formula, the time for spin flip is inversely proportional to the intensity of the magnetic field in the x-y plane. Not only T_1 depends on the magnitude of the pulses along the (x-y) axes but also on their sign. Pulses of opposite signs will delay the time for spin flip events and increase the time scale T_1.

The short magnetic pulses due to collisional interactions can also lead to small variation in the z-component of the magnetic field felt by each spin. This leads to a spread in the Larmor frequency for each individual spin and a resulting spread in the x and y components of the spinors on the Bloch sphere. For the observer in a frame rotating at the Larmor frequency, the spinors will no longer appear frozen in time and the relative positions of the spinors will change rapidly. The timescale over which this dephasing between spinors takes place can be very short since the sign of the z-component of the magnetic pulses can be either positive or negative. The timescale T_2 for phase randomization in the (x,y) plane is called the *transverse relation time*. Contrary to the case of the timescale T_1, magnetic pulses along the z-axis with opposite signs will lead to a more rapid spread in the x and y components of the spinors, hence to a shorter timescale T_2.

The general arguments above suggest that the timescales T_1 and T_2 are typically such that T_2 will be much smaller than T_1*. There are many other scattering mechanisms which can contribute to both relaxation times and the arguments above are not restricted to the spinors associated with atomic species of a gas in a cavity. In fact, the introduction of the timescale T_1 and

*This is actually not always true and there are situations when the reverse may be true. What is true has to be ascertained on a case-by-case basis.

T_2 was originally developed by Bloch to support the experimental effort of Rabi and his group in the earlier development of nuclear magnetic resonant experiments [1, 2, 3].

Bloch proposed that to describe phenomenologically the interaction of spinors with the environment, Equations (5.82 - 5.84) should be modified to include relaxation mechanisms which eventually help a collection of spinors reach their thermodynamical state of equilibrium. In this model, the new set of equations needed to characterize the relaxation of the generalized Bloch vector is given by the **Bloch equations**:

$$\frac{d\nu_x}{dt} = -\gamma(B_y\nu_z - \nu_y B_z) - \nu_x {T_2}^{-1}, \tag{5.88}$$

$$\frac{d\nu_y}{dt} = -\gamma(B_z v_x - v_z B_x) - \nu_y {T_2}^{-1} \tag{5.89}$$

and

$$\frac{d\nu_z}{dt} = -\gamma(B_x v_y - v_x B_y) - \nu_z {T_1}^{-1}, \tag{5.90}$$

where the components (B_x, B_y, B_z) are the components of the externally applied magnetic field whereas the last terms have been added to account phenomenologically for the rapidly varying magnetic field pulses arising from collisions among atoms or other sources representing additional interactions with the environment (like collisions of each individual atom with the walls of a cavity which could also be modeled as an effective magnetic field acting on each individual spin).

If we neglect the first terms on the right hand side of the Bloch equations, the solutions of the latter become

$$\nu_x(t) = \nu_x(0)\exp(-t/T_2), \tag{5.91}$$

$$\nu_y(t) = \nu_y(0)\exp(-t/T_2), \tag{5.92}$$

$$\nu_z(t) = (\nu_z(0) - \nu_{z0})\exp(-t/T_1) + \nu_{z0}, \tag{5.93}$$

which show that the *transverse* (x and y) components of the generalized Bloch vector in the equatorial plane decay exponentially to zero on a timescale T_2, whereas the *longitudinal* z-component decays exponentially to its equilibrium value ν_{z0} with time constant T_1. This is why T_1 is labeled the "longitudinal" relaxation time and T_2 is labeled the "transverse" relaxation time.

The Bloch equations have been used extensively to describe a variety of physical systems [1]. Typically, $T_2 < T_1$ when the effects of the environment can be described as a small perturbation to the effects of an externally applied magnetic field. In the next section, we use numerical simulations based on a Monte-Carlo approach to describe the time evolution of an ensemble of spinors submitted to a macroscopic external magnetic field but also additional rapidly varying magnetic pulses with intensity which can be much larger than the externally applied field. Even in that case, the relaxation of an ensemble of

spinors can be fitted to simple analytical functions that are solutions to the Bloch equations with phenomenological time constants for which the timescale T_1 can be smaller than T_2. This shows the generality of the Bloch equations in describing the effects of the environment on an ensemble of spins. This is a very valuable tool that underpins the importance of various (and sometimes competing) physical phenomena contributing to spin relaxation.

Numerical Examples

To illustrate the concept of density matrix for a mixed state and the corresponding time evolution of the generalized Bloch vector, a Monte-Carlo approach is used to study the influence of a constant magnetic field B_0 applied along the z-axis and a random magnetic field pulse applied along the x-axis on an ensemble of spinors, each of which is characterized by a Bloch vector at $t = 0$ given by

$$(n_x(0), n_y(0), n_z(0)) = (\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}). \tag{5.94}$$

The random magnetic field along the x-axis is assumed to be different for each spin in the ensemble and represents the effective magnetic field acting on each spin due to its interaction with the environment (which could be due to the influence of an effective magnetic field caused by the presence of other spins during collisions, or the influence of a time varying spin-orbit interaction, and any other type of interaction which can be modeled as a randomly fluctuating magnetic field). Since the spins are at different locations, the effective magnetic field they will feel will vary from spin to spin and, for simplicity, we assume that this effective magnetic field has only a component along the x-direction. Furthermore, we assume that the randomly fluctuating magnetic field changes over a typical time scale Δt whose value is comparable to the period of the Larmor precession of each spin due to the magnetic field B_0 along the z-axis. During each pulse, the field intensity along the x-axis is assumed to be constant but changed randomly between $\pm {B_1}^{max}$, as shown in Fig. 5.2.

Case 1: ${B_1}^{max}/B_0 = 10.0$ and $\omega_1 = -\gamma {B_1}^{max}$ is positive

Figures 5.3 through 5.5 show the time dependence of the average (over an ensemble of 1,000 spins) of the x, y and z components of the Bloch vector. The time interval Δt is selected such that $T/\Delta t$ is equal to 10, 50 and 100, respectively, where T is the period associated with the Larmor precession due to the magnetic field along the z-axis alone. All three components $(\bar{n}_x, \bar{n}_y, \bar{n}_z)$ rapidly decay towards 0, i.e, the Bloch vector representing the ensemble of spins moves quickly from the surface of the Bloch sphere towards the center of the Bloch ball. Since ${B_1}^{max} = 10B_0$, the precession period of the Bloch

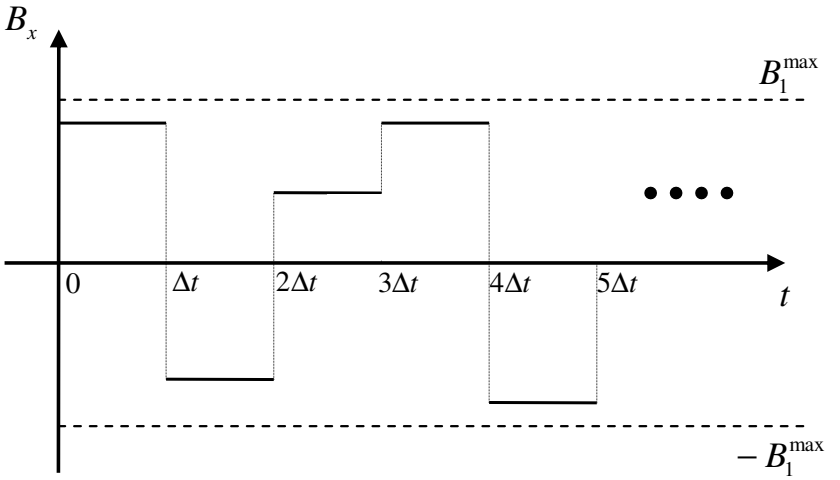


FIGURE 5.2
Random magnetic pulse sequence applied along the x-axis where the field intensity is changed randomly and uniformly between $\pm {B_1}^{max}$ at every time interval Δt.

vector due to B_1 alone around the x-axis is ten times smaller than the period for Larmor precession around the z-axis. As a result, $\bar{n}_y$ and $\bar{n}_z$ decay towards 0 very quickly. The decay is more rapid as the magnetic field pulse is changed less frequently, i.e., for smaller values of $T/\Delta t$ since, when the Larmor precession around the x-axis changes sign less frequently, $\bar{n}_y$ and $\bar{n}_z$ will average out to zero faster.

Consider $\bar{n}_x$. If Δt is very large, i.e., $T/\Delta t$ converges towards 0, the component B_x will not change very frequently and the generalized Bloch vector will rotate on a cone with principal axis along (B_1, 0., B_0) with B_1 constant. The average component $\bar{n}_x$ will therefore have mostly positive values. It is only after flipping the sign of B_x more frequently in the negative direction that $\bar{n}_x$ can pick up negative components. As a result, $\bar{n}_x$ decays faster towards zero as $T/\Delta t$ increases.

As the pulse rate increases, both $\bar{n}_x$ and $\bar{n}_y$ show some oscillatory behavior with decaying amplitude, as illustrated in Figures 5.6 and 5.7. The dashed lines in Figures 5.6 and 5.7 are fits to the Monte-Carlo simulation results based on analytical expressions of the form

$$\bar{n}_x = A_x cos(\omega t + \theta_x) e^{-\frac{\alpha_x t}{T}}, \tag{5.95}$$

and

$$\bar{n}_y = A_y cos(\omega t + \theta_y) e^{-\frac{\alpha_y t}{T}}. \tag{5.96}$$

The values of the parameters (A_x, A_y), (θ_x, θ_y) and (α_x, α_y) are listed in Tables 5.1 and 5.2 as a function of $N = T/\Delta t$. Within the statistical fluctuations of the Mote-Carlo results, the amplitudes A_x and A_y are basically equal to each other and the difference between θ_x and θ_y is equal to $\pi/2$. Furthermore, α_x and α_y are essentially identical. This implies, that in the (x, y) plane, the $(\bar{n}_x, \bar{n}_y)$ vector rotates uniformly with an angular frequency ω and the amplitude of the vector decays progressively towards zero. As N increases, the value of ω (in units of T^{-1}) converges towards 2π, i.e., approaches the Larmor frequency due to B_0 alone.

On the other hand, the decay of $\bar{n}_z$ is monotonic and well approximated by an exponential, as shown in Figure 5.8, i.e.,

$$\bar{n}_z = \frac{1}{\sqrt{3}} e^{-\frac{\alpha_z t}{T}}. \tag{5.97}$$

The absence of oscillations in $\bar{n}_z$ is due to the fact that frequent changes in sign of B_x as $T/\Delta t$ increases prevent a completion of a full Larmor precession around the $\hat{x}$ axis as positive and negative pulses offset each other's influence on $\bar{n}_z$. This leads to the gradual decay of $\bar{n}_z$ towards zero.

The decay rates of the amplitudes of $\bar{n}_x$ and $\bar{n}_y$ are identical and lower than the decay rate of $\bar{n}_z$ which is listed in the last column of Table 5.2. This implies that

$$T_1 = \frac{T}{\alpha_z} < T_2 = T_x = T_y = \frac{T}{\alpha_x} = \frac{T}{\alpha_y}, \tag{5.98}$$

i.e., the longitudinal relaxation time is smaller than the transverse relaxation time in the numerical simulations described above.

TABLE 5.1
Fitting parameters for x-component of generalized Bloch vector

$N = T/\Delta t$	A_x	ω	θ_x	α_x
100	0.621	5.766	0.433	3.279
200	0.692	6.106	0.638	1.609
300	0.790	6.089	0.722	1.249

Case 2: ${B_1}^{max}/B_0 = 1.0$ and $\omega_1 = -\gamma {B_1}^{max}$ is positive

Figures 5.9 through 5.11 show the time dependence of the average (over an ensemble of 1,000 spins) of the x, y, and z components of the Bloch vector for the case where the ratio ${B_1}^{max}/B_0$ is equal to 1, i.e., ten times smaller than in case 1. In this case, the effect of the magnetic field along the z-axis is more prominent and the oscillations in the $\bar{n}_x$ and $\bar{n}_y$ components of the generalized Bloch vector take longer to decay, as can be seen by comparing

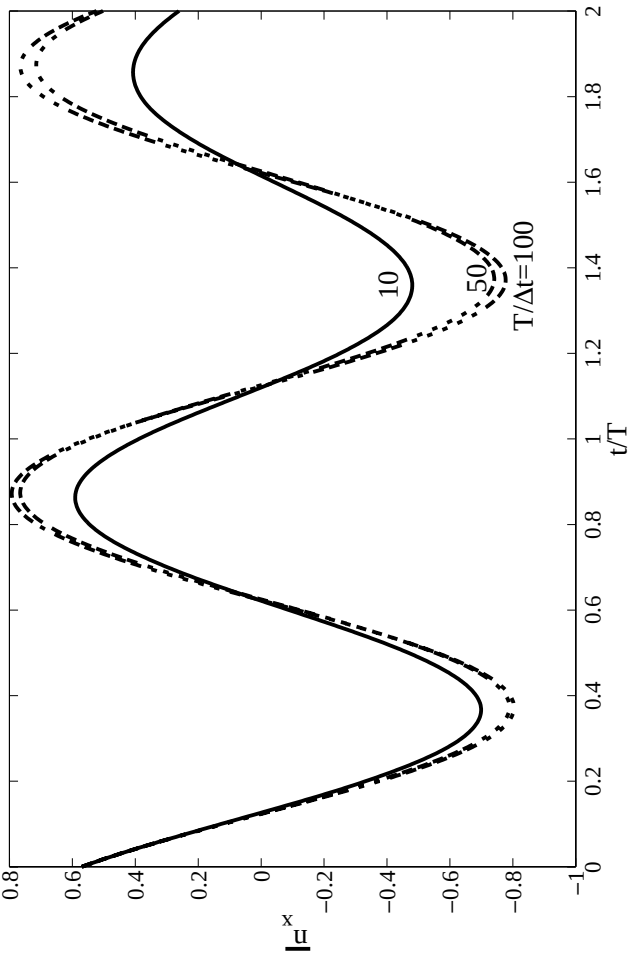


FIGURE 5.3
Time dependence of the average (over 1000 configurations) of the x-component of the generalized Bloch vector associated with an ensemble of qubits originally at $(n_x(0), n_y(0), n_z(0)) = (\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}})$ and submitted to the action of a constant magnetic field along the z-axis and a magnetic field square pulse along the x-axis where the intensities of the two magnetic fields are selected such that ${B_1}^{max}/B_0 = 10.0$ and $\omega_1 = -\gamma {B_1}^{max}$ is positive.

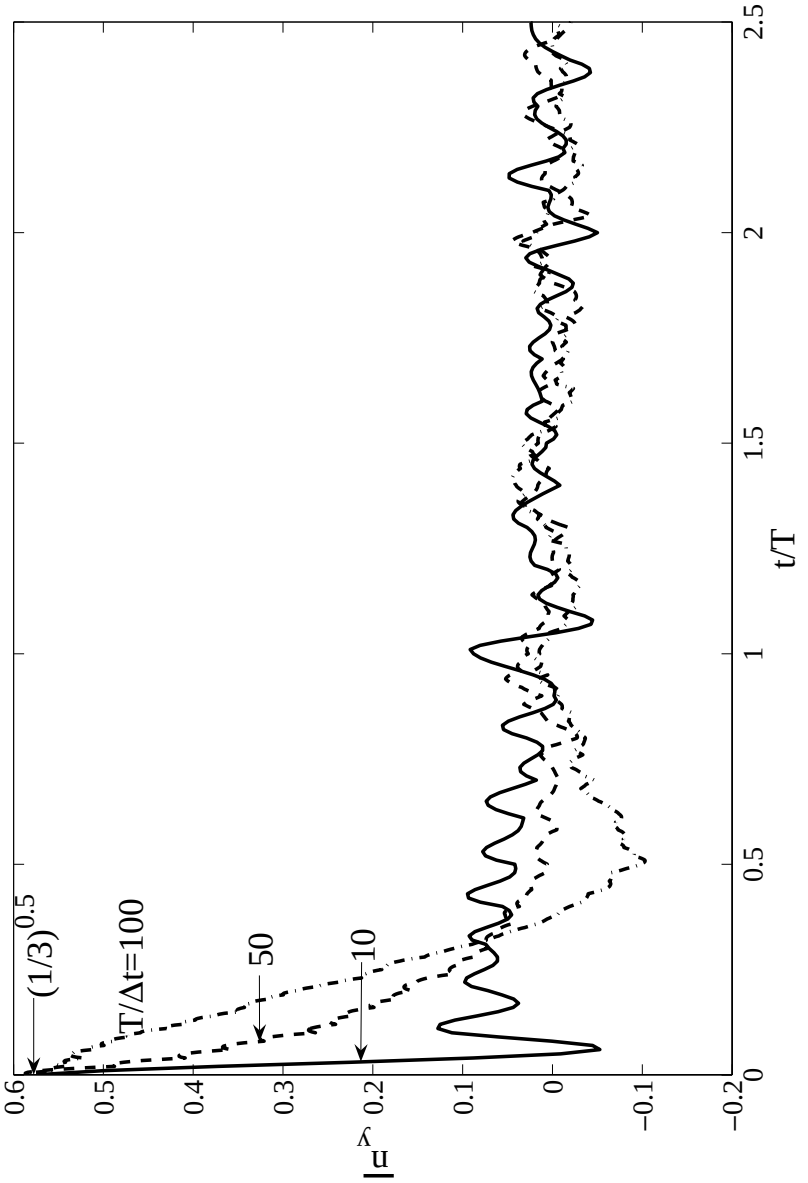


FIGURE 5.4
Same as Fig. 5.3 for the average component $\bar{n}_y(t)$ of the generalized Bloch vector associated with an ensemble of 1000 qubits located at $(n_x(0), n_y(0), n_z(0)) = (\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}})$ at time $t = 0$.

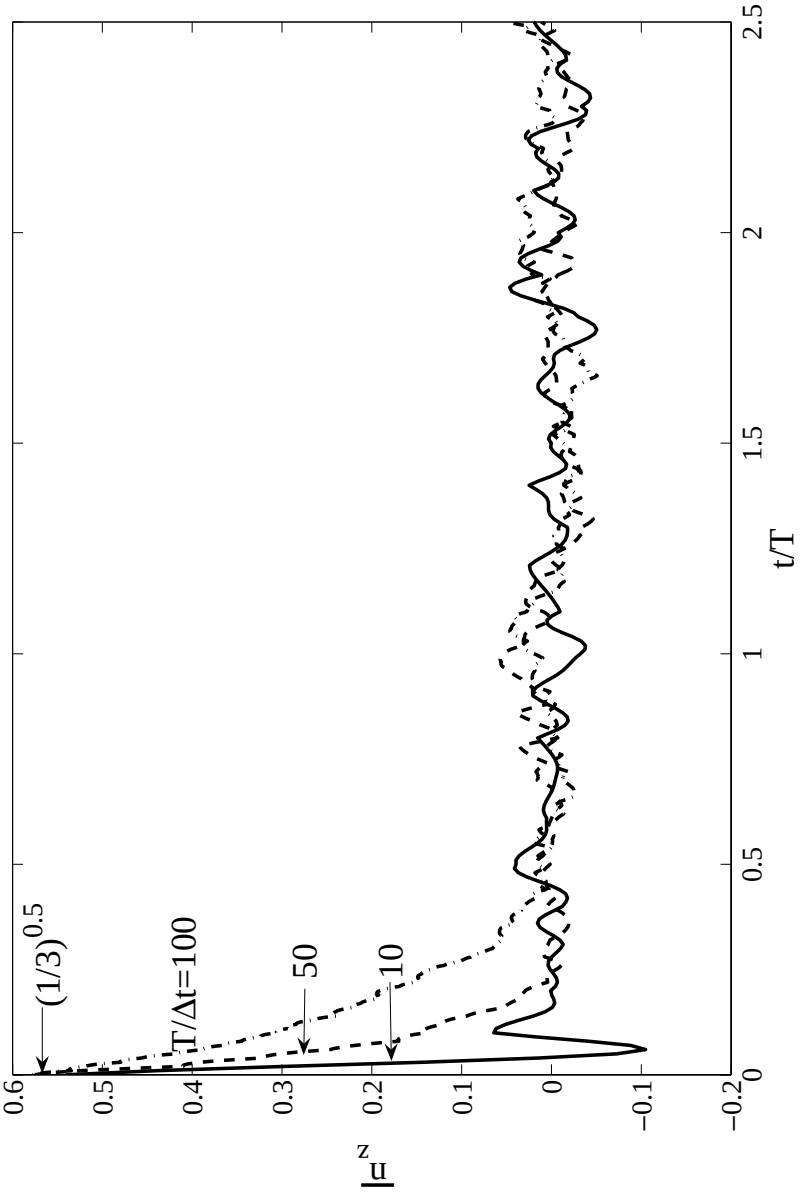


FIGURE 5.5
Same as Fig. 5.3 for the average component $\bar{n}_z(t)$ of the generalized Bloch vector associated with an ensemble of 1000 qubits located at $(n_x(0), n_y(0), n_z(0)) = (\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}})$ at time $t = 0$.

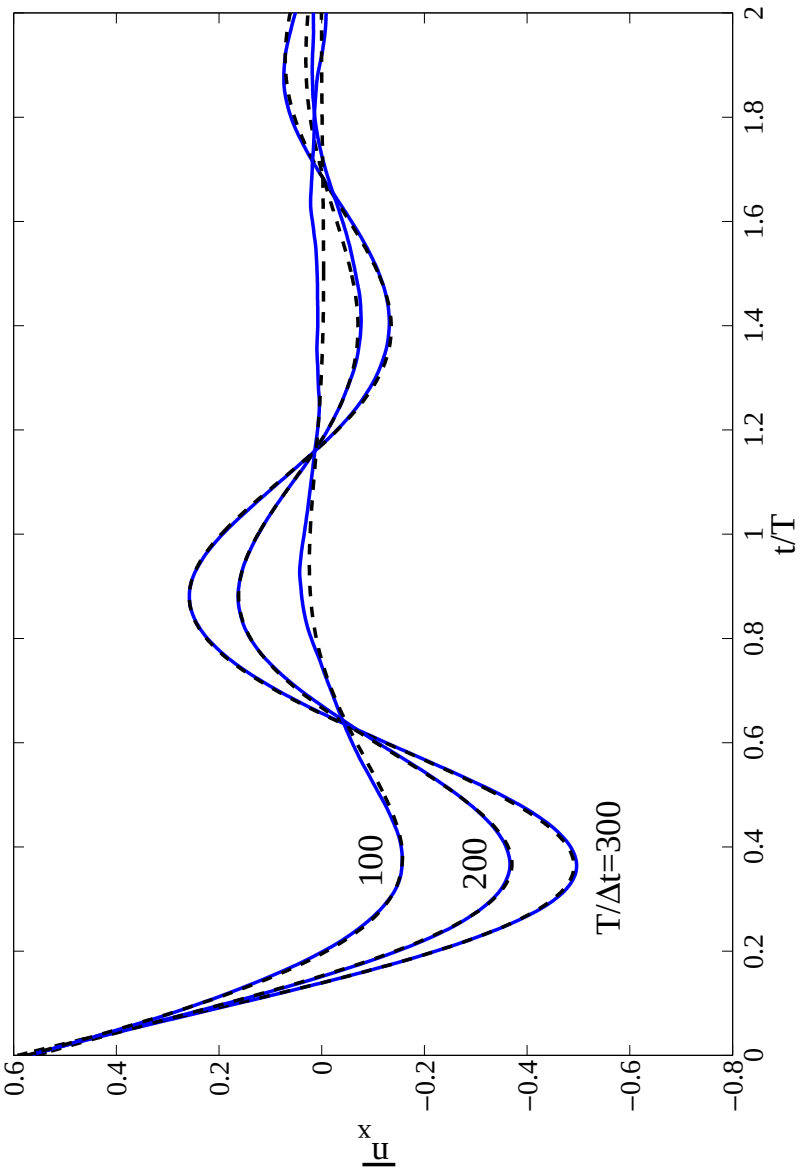


FIGURE 5.6
Same as Fig. 5.3 for small pulse, i.e., for larger values of $T/\Delta t$. The dashed curve is a fit to the Monte-Carlo simulations using the analytical expression given in Equation (5.95).

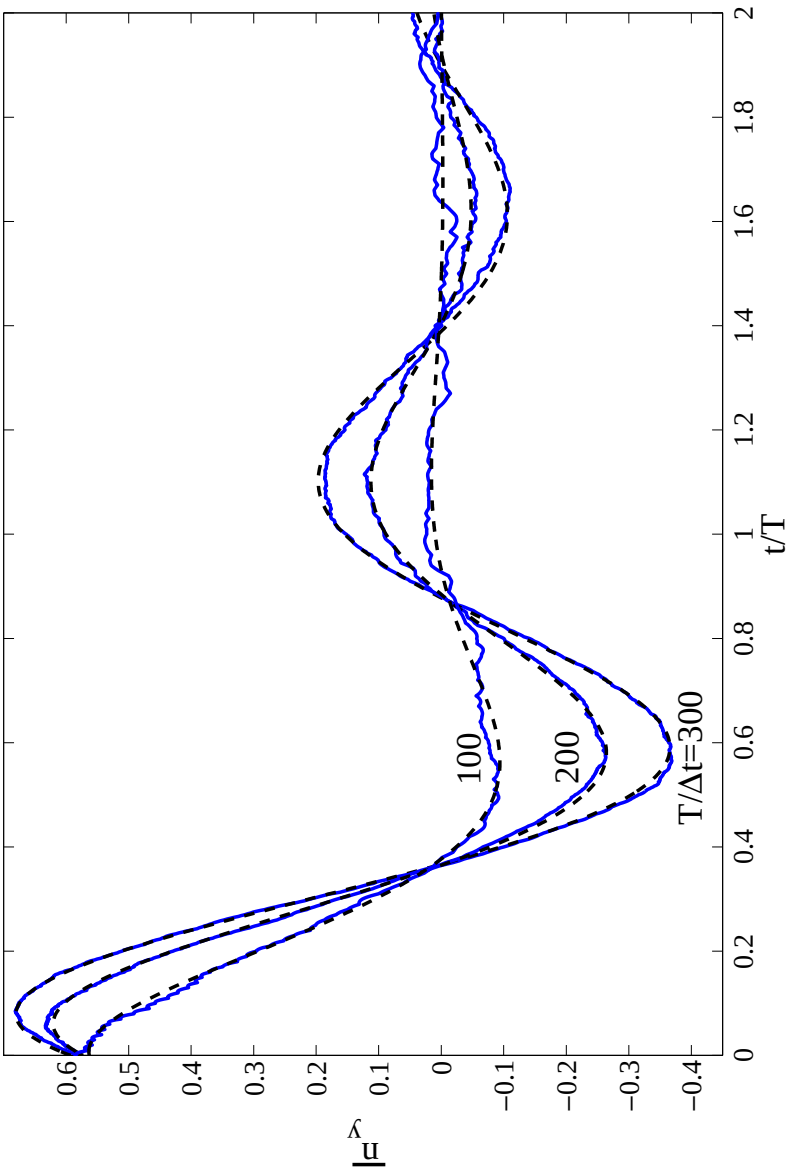


FIGURE 5.7
Same as Fig. 5.4 for small pulse, i.e., for larger values of $T/\Delta t$. The dashed curve is a fit to the Monte-Carlo simulations using the analytical expression given in Equation (5.96).

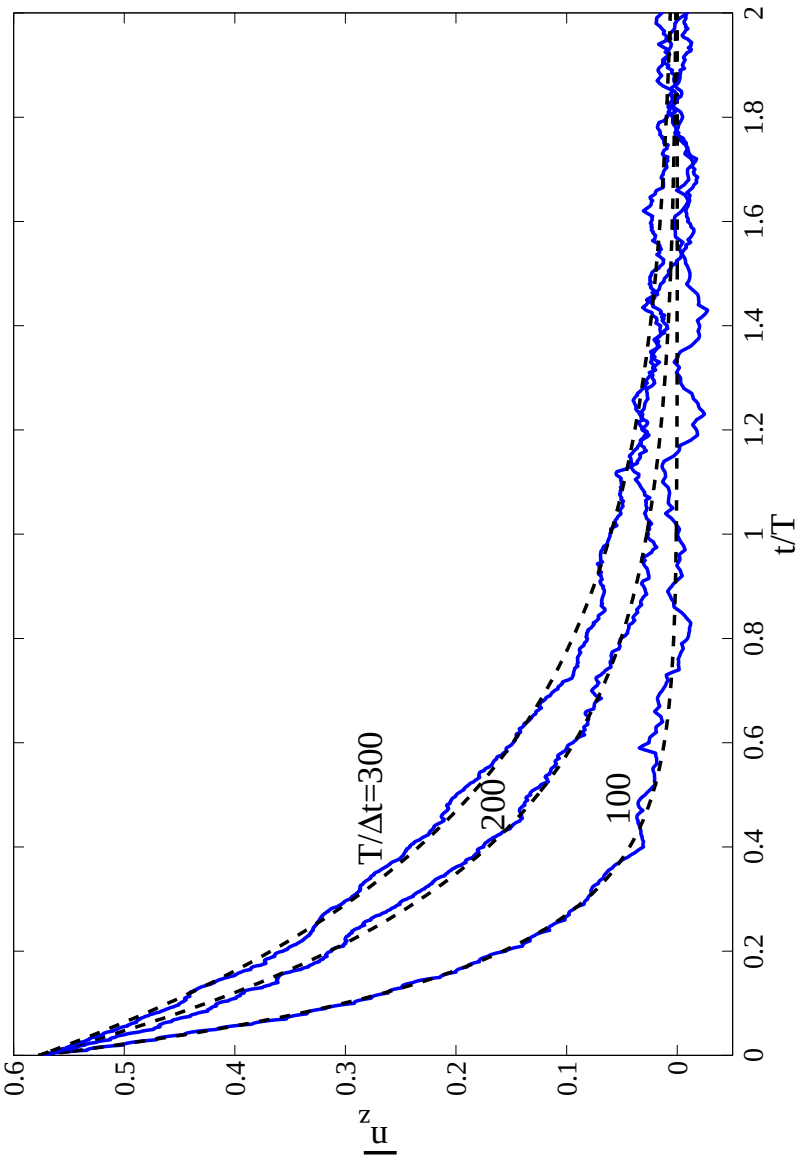


FIGURE 5.8
Same as Fig. 5.5 for small pulse, i.e., for larger values of $T/\Delta t$. The dashed curve is a fit to the Monte-Carlo simulations using the analytical expression given in Equation (5.97).

TABLE 5.2
Fitting parameters for y- and z-components of generalized Bloch vector

$N = T/\Delta t$	A_y	ω	θ_y	α_y	α_z
100	0.664	5.660	-0.563	3.226	6.49
200	0.711	6.045	-0.634	1.643	3.04
300	0.763	6.144	-0.682	1.213	2.26

Figures 5.9 and 5.10, and 5.3 and 5.4, respectively. Also, Figure 5.11 shows that $\bar{n}_z$ decays more slowly for the same value of $T/\Delta t$, when compared to the results displayed in Figure 5.5.

The numerical results described above show that the time evolution of the generalized Bloch vector associated with a collection of spinors seem to be in agreement with the phenomenological approach based on the Bloch equations. The values of the fitting parameters (relaxation times T_1 and T_2) depend strongly on the signal (amplitude and time variation) modeling the effects of the environment. The time T_1 and T_2 can also be a strong function of the external magnetic field B_0. For the case of spin transport through semiconductors, one needs to clearly identify the possible sources of time relaxation mechanisms which can affect T_1 and T_2. This subject is revisited in Chapter 8.

5.7 Problems

- **Problem 5.1**

  Starting with Equation (5.68), prove Equations (5.79) and (5.80) for the time evolution of the y and z components of the generalized Bloch vector.

  Solution

$$\frac{d\nu_y}{dt} = \frac{i\gamma}{4} Tr[\sigma_y\sigma_x\sigma_y(B_x\nu_y - \nu_x B_y) + \sigma_y\sigma_x\sigma_z(B_x\nu_z - \nu_x B_z) + \sigma_y\sigma_y\sigma_x(B_y\nu_x - \nu_y B_x)] + \frac{i\gamma}{4} Tr[\sigma_y\sigma_y\sigma_z(B_y\nu_z - \nu_y B_z) + \sigma_y\sigma_z\sigma_x(B_z\nu_x - \nu_z B_x) + \sigma_y\sigma_z\sigma_y(B_z\nu_y - \nu_z B_y)]. \quad (5.99)$$

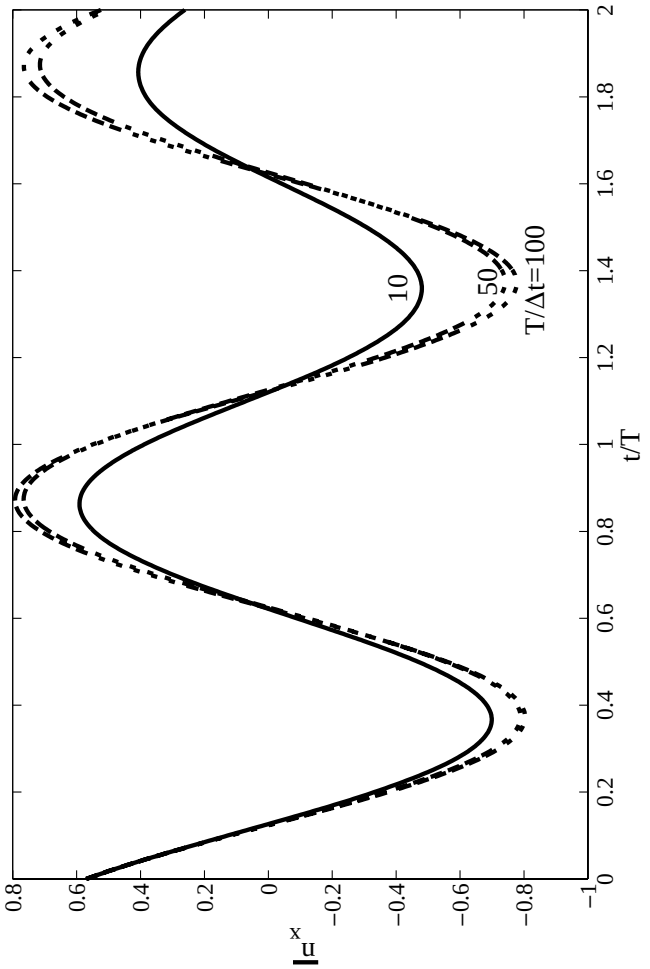


FIGURE 5.9
Time dependence of the average (over 1000 configurations) of the x-component of the generalized Bloch vector associated with an ensemble of qubits originally at $(n_x(0), n_y(0), n_z(0)) = (\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}})$ and submitted to the action of a constant magnetic field along the z-axis and a magnetic field square pulse along the x-axis where the intensities of the two magnetic fields are selected such that ${B_1}^{max}/B_0 = 1.0$ and $\omega_1 = -\gamma {B_1}^{max}$ is positive.

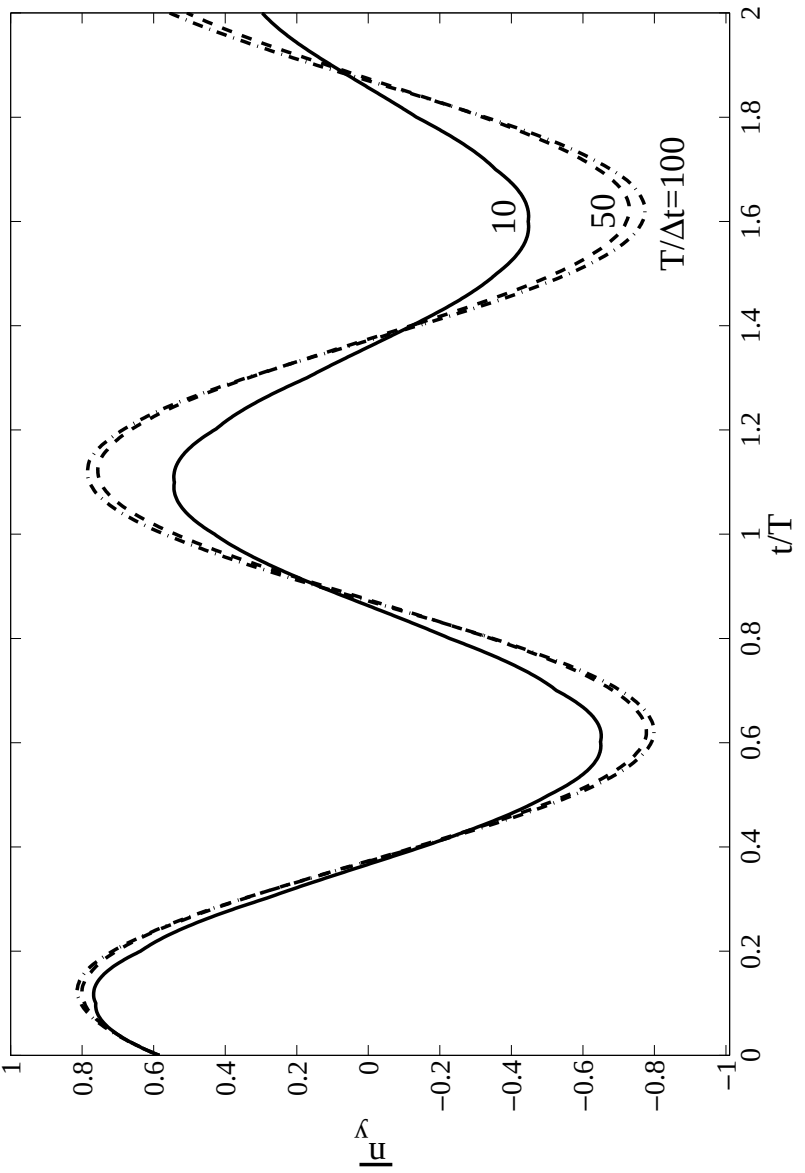


FIGURE 5.10
Same as Fig. 5.9 for the average component $\bar{n}_x(t)$ of the generalized Bloch vector associated with an ensemble of 1000 qubits located at $(n_x(0), n_y(0), n_z(0)) = (\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}})$ at time t = 0.

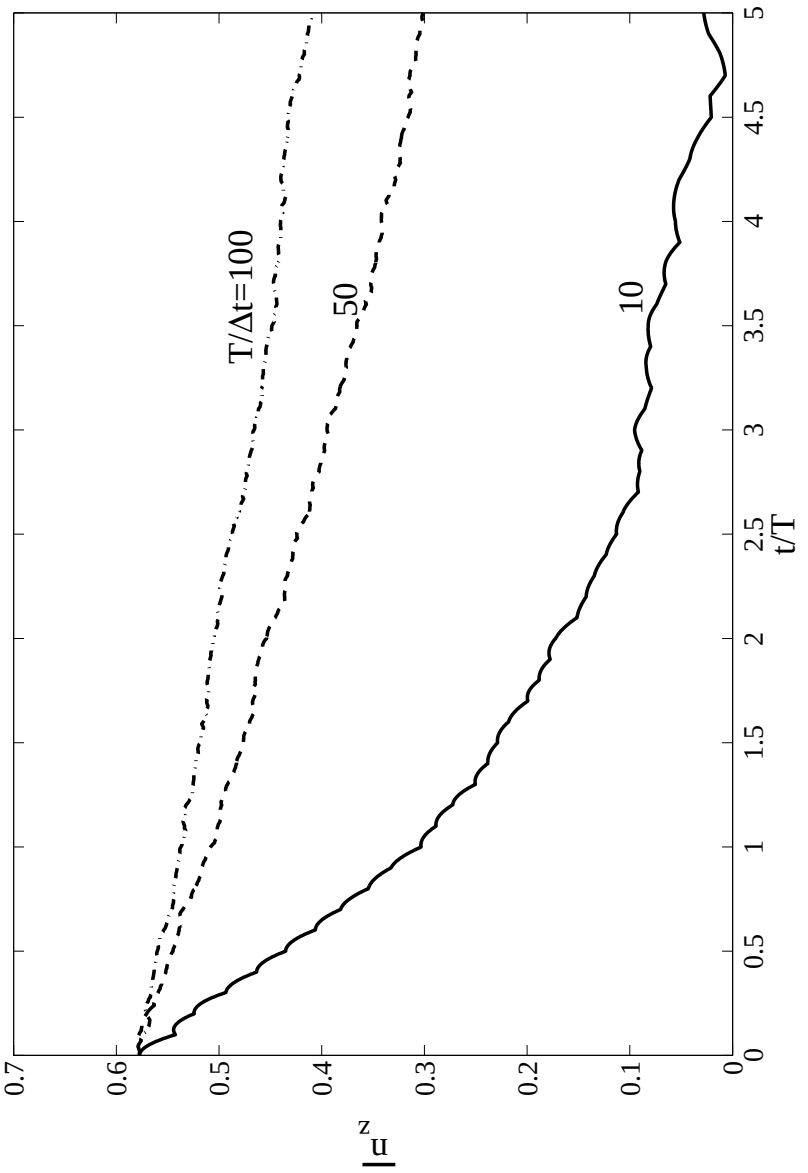


FIGURE 5.11
Same as Fig. 5.9 for the average component $\bar{n}_x(t)$ of the Bloch vector associated with a qubit located at $(n_x(0), n_y(0), n_z(0)) = (\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}})$ at time t = 0.

Using the cyclic property of the trace,

$$\begin{aligned}\frac{d\nu_y}{dt} = \frac{i\gamma}{4}Tr[&\sigma_y\sigma_y\sigma_x(B_x\nu_y - \nu_x B_y)\\ &+\sigma_y\sigma_x\sigma_z(B_x\nu_z - \nu_x B_z)\\ &+\sigma_y\sigma_y\sigma_x(B_y\nu_x - \nu_y B_x)]\\ &+\frac{i\gamma}{4}Tr[\sigma_y\sigma_y\sigma_z(B_y\nu_z - \nu_y B_z)\\ &+\sigma_y\sigma_z\sigma_x(B_z\nu_x - \nu_z B_x)\\ &+\sigma_y\sigma_y\sigma_z(B_z\nu_y - \nu_z B_y)].\end{aligned} \tag{5.100}$$

Since ${\sigma_y}^2 = 1$, we get

$$\begin{aligned}\frac{d\nu_y}{dt} = \frac{i\gamma}{4}Tr[&\sigma_x(B_x\nu_y - \nu_x B_y) + \sigma_y\sigma_x\sigma_z(B_x\nu_z - \nu_x B_z)\\ &+\sigma_x(B_y\nu_x - \nu_y B_x)]\\ +\frac{i\gamma}{4}Tr[&\sigma_z(B_y\nu_z - \nu_y B_z) + \sigma_y\sigma_z\sigma_x(B_z\nu_x - \nu_z B_x)\\ &+\sigma_z(B_z\nu_y - \nu_z B_y)].\end{aligned} \tag{5.101}$$

Hence,

$$\begin{aligned}\frac{d\nu_y}{dt} = &\frac{i\gamma}{4}Tr\left[\sigma_y\sigma_x\sigma_z(B_x\nu_z - \nu_x B_z)\right]\\ &+\frac{i\gamma}{4}Tr\left[\sigma_y\sigma_z\sigma_x(B_z\nu_x - \nu_z B_x)\right].\end{aligned} \tag{5.102}$$

Pulling out the term $(B_z\nu_x - \nu_z B_x)$, we get

$$\frac{d\nu_y}{dt} = \frac{i\gamma}{4}(B_z\nu_x - \nu_z B_x)Tr\left[\sigma_y\sigma_x\sigma_z - \sigma_y\sigma_z\sigma_x\right].$$

Using the identities $\sigma_y\sigma_x\sigma_z = -iI$ and $\sigma_y\sigma_z\sigma_x = iI$, we get $\sigma_y\sigma_x\sigma_z - \sigma_y\sigma_z\sigma_x = -2iI$, and therefore

$$\frac{d\nu_y}{dt} = -\gamma(B_z\nu_x - \nu_z B_x). \tag{5.103}$$

Similarly,

$$\begin{aligned}\frac{d\nu_z}{dt} = \frac{i\gamma}{4}Tr[&\sigma_z\sigma_x\sigma_y(B_x\nu_y - \nu_x B_y)\\ &+\sigma_z\sigma_x\sigma_z(B_x\nu_z - \nu_x B_z)\\ &+\sigma_z\sigma_y\sigma_x(B_y\nu_x - \nu_y B_x)]\end{aligned}$$

$$
\begin{aligned}
&+\frac{i\gamma}{4}Tr[\sigma_z\sigma_y\sigma_z(B_y\nu_z-\nu_yB_z)\\
&+\sigma_z\sigma_z\sigma_x(B_z\nu_x-\nu_zB_x)\\
&+\sigma_z\sigma_z\sigma_y(B_z\nu_y-\nu_zB_y)]. \qquad (5.104)
\end{aligned}
$$

Using the cyclic property of the trace once again

$$
\begin{aligned}
\frac{d\nu_z}{dt} = &\frac{i\gamma}{4}Tr[\sigma_z\sigma_x\sigma_y(B_x\nu_y-\nu_xB_y)\\
&+\sigma_z\sigma_z\sigma_x(B_x\nu_z-\nu_xB_z)\\
&+\sigma_z\sigma_y\sigma_x(B_y\nu_x-\nu_yB_x)]\\
&+\frac{i\gamma}{4}Tr[\sigma_z\sigma_z\sigma_y(B_y\nu_z-\nu_yB_z)\\
&+\sigma_z\sigma_z\sigma_x(B_z\nu_x-\nu_zB_x)\\
&+\sigma_z\sigma_z\sigma_y(B_z\nu_y-\nu_zB_y)]. \qquad (5.105)
\end{aligned}
$$

Since ${\sigma_z}^2 = 1$, we get

$$
\begin{aligned}
\frac{d\nu_z}{dt} = &\frac{i\gamma}{4}Tr[\sigma_z\sigma_x\sigma_y(B_x\nu_y-\nu_xB_y)\\
&+\sigma_x(B_x\nu_z-\nu_xB_z)\\
&+\sigma_z\sigma_y\sigma_x(B_y\nu_x-\nu_yB_x)]\\
&+\frac{i\gamma}{4}Tr[\sigma_y(B_y\nu_z-\nu_yB_z)\\
&+\sigma_x(B_z\nu_x-\nu_zB_x)\\
&+\sigma_y(B_z\nu_y-\nu_zB_y)]. \qquad (5.106)
\end{aligned}
$$

Hence,

$$
\frac{d\nu_z}{dt} = \frac{i\gamma}{4}Tr\left[\sigma_z\sigma_x\sigma_y(B_x\nu_y-\nu_xB_y)+\sigma_z\sigma_y\sigma_x(B_y\nu_x-\nu_yB_x)\right]. \qquad (5.107)
$$

Pulling out the term $(B_x\nu_y - \nu_xB_y)$, we get

$$
\frac{d\nu_z}{dt} = \frac{i\gamma}{4}(B_x\nu_y-\nu_xB_y)Tr\left[\sigma_z\sigma_x\sigma_y-\sigma_z\sigma_y\sigma_x\right]. \qquad (5.108)
$$

Using the identities $\sigma_z\sigma_x\sigma_y = iI$ and $\sigma_z\sigma_y\sigma_x = -iI$, we get $\sigma_z\sigma_x\sigma_y - \sigma_z\sigma_y\sigma_x = 2iI$, leading to

$$
\frac{d\nu_z}{dt} = -\gamma(B_x\nu_y-\nu_xB_y). \qquad (5.109)
$$

- **Problem 5.2**

  In the ensuing problems, we show how the equation of motion for the density matrix can be used to re-derive Equation (4.1) in Chapter 4 which is the starting point of the discussion of the Rabi oscillations and the Rabi formula.

  Starting with Equation (5.24), establish the following relationship between the general expression for a spinor in a frame of reference rotating at a constant angular velocity ω around the z-axis (counterclockwise) and the expression for the same spinor expressed with respect to a fixed frame of reference:

$$|\psi^{'}\rangle = e^{i\gamma} e^{i\frac{\omega t}{2}\sigma_z} |\psi\rangle. \tag{5.110}$$

  What is the expression for the phase γ?

  Solution

  In the fixed frame of reference,

$$|\psi\rangle = |\xi_n^{\dagger}\rangle = \cos\frac{\theta}{2}|0\rangle + \sin\frac{\theta}{2} e^{i\phi}|1\rangle. \tag{5.111}$$

  In the rotating frame, the expression of the spinor is obtained from the previous one by replacing ϕ by $\phi - \omega t$. Hence,

$$|\psi^{'}\rangle = e^{i\frac{\omega t}{2}} \begin{bmatrix} e^{i\frac{\omega t}{2}} & 0 \\ 0 & e^{-i\frac{\omega t}{2}} \end{bmatrix} |\psi\rangle \tag{5.112}$$

  or, more succinctly,

$$|\psi^{'}\rangle = e^{i\frac{\omega t}{2}} e^{i\frac{\omega t}{2}\sigma_z} |\psi\rangle. \tag{5.113}$$

  Hence, the phase factor γ is Equation (5.111) is given by $\frac{\omega t}{2}$.

- **Problem 5.3**

  Show that the expression for the density matrix $\rho^{'}$ in the rotating frame of reference, i.e., associated with the spinor $|\psi^{'}\rangle$, is related to the density matrix ρ associated with the spinor $|\psi\rangle$ in the fixed reference frame as follows

$$\rho^{'} = U_z \rho U_z^{\dagger}, \tag{5.114}$$

  where U_z is the unitary matrix

$$U_z = e^{i\omega\sigma_z t/2}. \tag{5.115}$$

- **Problem 5.4**

  Using the results of the previous problem and Equation (5.59) for the time evolution of the density matrix in the fixed frame of reference, show

that the time evolution of the density matrix in the rotating frame of reference is given by

$$\frac{d\rho^{'}}{dt} = \frac{-i}{\hbar}\left[H^{'}, \rho^{'}\right], \tag{5.116}$$

where the Hamiltonian $H^{'}$ in the rotating frame is related to the Hamiltonian H in the fixed frame as follows:

$$H^{'} = U_z H {U_z}^{-1} - i\hbar U_z \frac{d{U_z}^{-1}}{dt}. \tag{5.117}$$

Note that with the use of Equation (5.115), the last equation can be rewritten as follows

$$H' = U_z H {U_z}^{-1} - \frac{\hbar\omega}{2}\sigma_z. \tag{5.118}$$

- **Problem 5.5**

  For the case where H is the Hamiltonian of a spin-1/2 particle in the presence of a constant magnetic field along the z-axis and a rotating magnetic field with intensity B_1 rotating with a constant angular velocity in the (x,y) plane and located along the x-axis at $t = 0$, we have

$$H = \omega_L S_z B_0 - \gamma B_1 (cos(\omega t) S_x + sin(\omega t) S_y). \tag{5.119}$$

  Show that the explicit form of the Hamiltonian $H^{'}$ is given by

$$H^{'} = \frac{1}{2}\hbar(\Delta\sigma_z - \gamma b_1 \sigma_x), \tag{5.120}$$

  where $\Delta = \omega_L - \omega$ and $\omega_L = -\gamma B_0$.

- **Problem 5.6**

  Since the equation governing the time evolution of the density matrix in the rotating frame is similar to the one derived in the fixed frame of reference, the equations of motion of the components of the Bloch vector in the rotating frame are the same as in the fixed frame and given by Equation (5.66), where all quantities (except for the Pauli matrices) must be primed.

  Neglecting the primes for simplicity, the equations of evolution of each component of the Bloch vector in the rotating frame is therefore given by

$$\frac{d\nu_i}{dt} = -\tfrac{i}{4} Tr\left[\sigma_i \left[\Delta\sigma_z - \gamma B_1 \sigma_x, (1 + \nu_x\sigma_x + \nu_y\sigma_y + \nu_z\sigma_z)\right]\right]. \tag{5.121}$$

Starting with the expression for the Hamiltonian in the rotating frame of reference, show that the equations of the time evolution of each component of the Bloch vector in the rotating frame is given in a vectorial form by

$$\frac{d}{dt}\vec{\nu} = \gamma\vec{\nu} \times \vec{B_{eff}}, \tag{5.122}$$

where

$$\vec{B_{eff}} = \left(B_0 + \frac{\omega}{\gamma}\right)\hat{z} + B_1\hat{x}. \tag{5.123}$$

Solution

Starting with Equation (5.121), we get

$$\begin{aligned}\frac{d\nu_i}{dt} = &-\tfrac{i}{4}Tr\left[\sigma_i(\Delta\sigma_z(\nu_x\sigma_x + \nu_y\sigma_y) - \gamma B_1\sigma_x(\nu_y\sigma_y + \nu_z\sigma_z))\right] \\ &+\tfrac{i}{4}Tr\left[\sigma_i(\Delta(\nu_x\sigma_x + \nu_y\sigma_y)\sigma_z - \gamma B_1(\nu_y\sigma_y + \nu_z\sigma_z)\sigma_x)\right].\end{aligned} \tag{5.124}$$

For the x-component, we obtain

$$\begin{aligned}\frac{d\nu_x}{dt} = &-\tfrac{i}{4}Tr\left[\Delta\sigma_x\sigma_z(\nu_x\sigma_x + \nu_y\sigma_y) - \gamma B_1\sigma_x\sigma_x(\nu_y\sigma_y + \nu_z\sigma_z))\right] \\ &+\tfrac{i}{4}Tr\left[\Delta(\nu_x\sigma_x\sigma_x + \nu_y\sigma_x\sigma_y)\sigma_z - \gamma B_1(\nu_y\sigma_x\sigma_y + \nu_z\sigma_x\sigma_z)\sigma_x)\right],\end{aligned} \tag{5.125}$$

$$\begin{aligned}\frac{d\nu_x}{dt} = &-\tfrac{i}{4}Tr\left[\Delta\nu_x\sigma_x\sigma_z\sigma_x + \Delta\nu_y\sigma_x\sigma_z\sigma_y - \gamma B_1(\nu_y\sigma_y + \nu_z\sigma_z))\right] \\ &+\tfrac{i}{4}Tr\left[\Delta(\nu_x\sigma_x\sigma_x + \nu_y\sigma_x\sigma_y)\sigma_z - \gamma B_1(\nu_y\sigma_x\sigma_y + \nu_z\sigma_x\sigma_z)\sigma_x)\right],\end{aligned} \tag{5.126}$$

$$\begin{aligned}\frac{d\nu_x}{dt} = &-\tfrac{i}{4}Tr\left[-\Delta\nu_x\sigma_z - i\Delta\nu_y I - \gamma B_1(\nu_y\sigma_y + \nu_z\sigma_z))\right] \\ &+\tfrac{i}{4}Tr\left[\Delta(\nu_x\sigma_z + i\nu_y I) - \gamma B_1(\nu_y\sigma_x\sigma_y\sigma_x + \nu_z\sigma_x\sigma_z\sigma_x))\right],\end{aligned} \tag{5.127}$$

$$\begin{aligned}\frac{d\nu_x}{dt} = &-\tfrac{i}{4}Tr\left[-\Delta\nu_x\sigma_z - i\Delta\nu_y I - \gamma B_1(\nu_y\sigma_y + \nu_z\sigma_z))\right] \\ &+\tfrac{i}{4}Tr\left[\Delta(\nu_x\sigma_z + i\nu_y I) - \gamma B_1(-\nu_y\sigma_y - \nu_z\sigma_z))\right],\end{aligned} \tag{5.128}$$

$$\frac{d\nu_x}{dt} = -\tfrac{i}{4}Tr\left[-\Delta\nu_x\sigma_z - i\Delta\nu_y I - \gamma B_1\nu_y\sigma_y - \gamma b_1\nu_z\sigma_z)\right] + \tfrac{i}{4}Tr\left[\Delta\nu_x\sigma_z + i\Delta\nu_y I + \gamma B_1\nu_y\sigma_y + \gamma b_1\nu_z\sigma_z\right], \tag{5.129}$$

which reduces to

$$\frac{d\nu_x}{dt} = \frac{i}{4}Tr\left[i\Delta\nu_y I\right] + \frac{i}{4}Tr\left[i\Delta\nu_y I\right], \tag{5.130}$$

and hence,

$$\frac{d\nu_x}{dt} = -\Delta\nu_y. \tag{5.131}$$

Similarly, the y-component of Equation (5.121) leads to

$$\frac{d\nu_y}{dt} = -\tfrac{i}{4}Tr\left[\sigma_y(\Delta\sigma_z(\nu_x\sigma_x + \nu_y\sigma_y) - \gamma b_1\sigma_x(\nu_y\sigma_y + \nu_z\sigma_z))\right] + \tfrac{i}{4}Tr\left[\sigma_y(\Delta(\nu_x\sigma_x + \nu_y\sigma_y)\sigma_z - \gamma B_1(\nu_y\sigma_y + \nu_z\sigma_z)\sigma_x)\right], \tag{5.132}$$

$$\frac{d\nu_y}{dt} = -\tfrac{i}{4}Tr\left[\Delta\sigma_y\sigma_z(\nu_x\sigma_x + \nu_y\sigma_y) - \gamma B_1\sigma_y\sigma_x(\nu_y\sigma_y + \nu_z\sigma_z))\right] + \tfrac{i}{4}Tr\left[\Delta\sigma_y(\nu_x\sigma_x + \nu_y\sigma_y)\sigma_z - \gamma B_1\sigma_y(\nu_y\sigma_y + \nu_z\sigma_z)\sigma_x)\right], \tag{5.133}$$

$$\frac{d\nu_y}{dt} = -\tfrac{i}{4}Tr\left[\nu_x\Delta\sigma_y\sigma_z\sigma_x + \nu_y\Delta\sigma_y\sigma_z\sigma_y - \nu_y\gamma B_1\sigma_y\sigma_x\sigma_y + \nu_z\gamma B_1\sigma_y\sigma_x\sigma_z\right] + \tfrac{i}{4}Tr\left[\nu_x\Delta\sigma_y\sigma_x\sigma_z + \nu_y\Delta\sigma_y\sigma_y\sigma_z - \gamma B_1\sigma_y\nu_y\sigma_y\sigma_x - \nu_z\gamma B_1\sigma_y\sigma_z\sigma_x\right], \tag{5.134}$$

$$\frac{d\nu_y}{dt} = -\tfrac{i}{4}Tr\left[\nu_x\Delta\sigma_y\sigma_z\sigma_x + \nu_y\Delta\sigma_y\sigma_y\sigma_z - \nu_y\gamma B_1\sigma_y\sigma_y\sigma_x + \nu_z\gamma B_1\sigma_y\sigma_x\sigma_z\right] + \tfrac{i}{4}Tr\left[\nu_x\Delta\sigma_y\sigma_x\sigma_z + \nu_y\Delta\sigma_y\sigma_y\sigma_z - \gamma B_1\nu_y\sigma_y\sigma_y\sigma_x - \nu_z\gamma B_1\sigma_y\sigma_z\sigma_x\right], \tag{5.135}$$

$$\frac{d\nu_y}{dt} = -\frac{i}{4}Tr\left[\nu_x\Delta\sigma_y\sigma_z\sigma_x + \nu_y\Delta\sigma_z - \nu_y\gamma b_1\sigma_x + \nu_z\gamma B_1\sigma_y\sigma_x\sigma_z\right] + \frac{i}{4}Tr\left[\nu_x\Delta\sigma_y\sigma_x\sigma_z + \nu_y\Delta\sigma_z - \gamma b_1\nu_y\sigma_x - \nu_z\gamma B_1\sigma_y\sigma_z\sigma_x\right], \tag{5.136}$$

$$\frac{d\nu_y}{dt} = -\frac{i}{4}Tr\left[\nu_x\Delta\sigma_y\sigma_z\sigma_x + \nu_z\gamma b_1\sigma_y\sigma_x\sigma_z\right] + \frac{i}{4}Tr\left[\nu_x\Delta\sigma_y\sigma_x\sigma_z - \nu_z\gamma b_1\sigma_y\sigma_z\sigma_x\right], \tag{5.137}$$

and eventually,

$$\frac{d\nu_y}{dt} = \Delta\nu_x + \frac{i\gamma B_1\nu_z}{4}Tr\left[\sigma_y\sigma_x\sigma_z - \sigma_y\sigma_z\sigma_x\right] = \Delta\nu_x + \gamma B_1\nu_z \tag{5.138}$$

- Finally, the z-component of Equation (5.121) is given by

$$\frac{d\nu_z}{dt} = -\frac{i}{4}Tr\left[\sigma_z(\Delta\sigma_z(\nu_x\sigma_x + \nu_y\sigma_y) - \gamma B_1\sigma_x(\nu_y\sigma_y + \nu_z\sigma_z))\right] + \frac{i}{4}Tr\left[\sigma_z(\Delta(\nu_x\sigma_x + \nu_y\sigma_y)\sigma_z - \gamma B_1(\nu_y\sigma_y + \nu_z\sigma_z)\sigma_x)\right], \tag{5.139}$$

$$\frac{d\nu_z}{dt} = -\frac{i}{4}Tr\left[\Delta\sigma_z\sigma_z(\nu_x\sigma_x + \nu_y\sigma_y) - \gamma B_1\sigma_z\sigma_x(\nu_y\sigma_y + \nu_z\sigma_z))\right] + \frac{i}{4}Tr\left[\nu_x\Delta\sigma_z\sigma_x\sigma_z + \nu_y\Delta\sigma_z\sigma_y\sigma_z - \gamma B_1\sigma_z\nu_y\sigma_y\sigma_x - \gamma B_1\nu_z\sigma_z\sigma_z\sigma_x\right], \tag{5.140}$$

$$\frac{d\nu_z}{dt} = -\frac{i}{4}Tr\left[\Delta\sigma_z\sigma_z\nu_x\sigma_x + \nu_y\Delta\sigma_z\sigma_z\sigma_y - \gamma B_1\nu_y\sigma_z\sigma_x\sigma_y - \gamma B_1\nu_z\sigma_z\sigma_x\sigma_z\right] + \frac{i}{4}Tr\left[\nu_x\Delta\sigma_z\sigma_x\sigma_z + \nu_y\Delta\sigma_z\sigma_y\sigma_z - \gamma B_1\nu_y\sigma_z\sigma_y\sigma_x - \gamma B_1\nu_z\sigma_z\sigma_z\sigma_x\right], \tag{5.141}$$

$$\frac{d\nu_z}{dt} = -\frac{i}{4}Tr\left[\Delta\nu_x\sigma_x + \nu_y\Delta\sigma_y - \gamma B_1\nu_y\sigma_z\sigma_x\sigma_y - \gamma B_1\nu_z\sigma_x\right] + \frac{i}{4}Tr\left[\nu_x\Delta\sigma_x + \nu_y\Delta\sigma_y - \gamma B_1\nu_y\sigma_z\sigma_y\sigma_x - \gamma B_1\nu_z\sigma_x\right], \tag{5.142}$$

$$\frac{d\nu_z}{dt} = -\frac{i}{4}Tr\left[-\gamma B_1 \nu_y \sigma_z \sigma_x \sigma_y\right] + \frac{i}{4}Tr\left[-\gamma B_1 \nu_y \sigma_z \sigma_y \sigma_x\right] \quad (5.143)$$

leading to

$$\frac{d\nu_z}{dt} = -\frac{i}{4}Tr\left[\sigma_z(-\gamma B_1 \sigma_x \nu_y \sigma_y) - \sigma_z(-\gamma b_1 \nu_y \sigma_y \sigma_x)\right] = -\gamma B_1 \nu_y. \quad (5.144)$$

Equations (5.131),(5.138), and (5.144) can be written in the more compact vectorial form (5.122).

- Starting with Equation (5.122), show that, in the frame rotating at the Larmor frequency, the effective magnetic field experienced by the spinor is a constant magnetic field of amplitude B_1 along the x-axis.

Solution

Starting with Equations (5.122-5.123) which describes the time evolution of the generalized Bloch vector in the rotating frame:

$$\vec{B_{eff}} = \left(B_0 + \frac{\omega}{\gamma}\right)\hat{z} + B_1\hat{x}, \quad (5.145)$$

the z-component of the effective magnetic field will be equal to zero when

$$B_0 + \frac{\omega}{\gamma} = 0, \quad (5.146)$$

i.e., when $\omega = \omega_L = -\gamma B_0$ (which is positive since, in our case, γ is negative). In this case the effective magnetic field is a constant magnetic field of intensity B_1 along the +x-axis and all the spinors will rotate around the x-axis of the rotating frame with an angular frequency equal to the Larmor frequency corresponding to the intensity B_1. If all the spinors start at the north pole, they will all reach the south pole simultaneously if the effects of relaxation due to interactions with the environment are neglected.

5.8 References

[1] K. Blum, *Density Matrix Theory and Applications*, (Plenum, New York, 1996).

[2] R. P. Feynman, *Statistical Mechanics*, Frontiers in Physics Lecture Note Series, (Addison-Wesley, Redwood City, California, 1972)

[3] C. Kittel, *Introduction to Solid State Physics*, 2nd edition, (John Wiley & Sons, London, 1961).

6

Spin Orbit Interaction

In the previous chapter, we introduced the relaxation times T_1 and T_2. One of the major mechanisms that determine T_1 and T_2 is *spin-orbit interaction*, because it causes spin relaxation. It is important to understand it in sufficient detail, not only for comprehending spin relaxation, but also since spin-orbit interaction is at the heart of many spin-based devices that are discussed in Chapter 12.

The American physicist Kronig who thought of spin as being associated with the self rotation of an electron about its own axis (recall Chapter 1) had thought of other things as well. The self rotation of a charged particle about its own axis will give rise to a magnetic moment and Kronig anticipated that the interaction between this magnetic moment and the orbital angular momentum of an electron in an atom could be derived from relativistic principles. This led to the discovery and formulation of spin-orbit interaction.

While orbiting around the nucleus, the negatively charged electron in an atom feels the electric field due to the positively charged nucleus. As a result, a magnetic field, that did not exist in the laboratory frame, will appear in the rest frame of the electron (see Problem 6.1) through a Lorentz transformation. According to Einstein's relativity theory, the flux density associated with this magnetic field is

$$\vec{\mathcal{B}} = \frac{\vec{\mathcal{E}} \times \vec{v}}{c^2\sqrt{1 - v^2/c^2}}, \tag{6.1}$$

where $\vec{\mathcal{E}}$ is the electric field seen by the electron, $\vec{v}$ is its orbital velocity and c is the speed of light in vacuum.

The above equation is actually not entirely correct. Thomas in a paper published in *Nature* in 1926 [1] had pointed out that the Lorentz transformation that we normally use to connect the electron's rest frame to the laboratory frame is inexact. The electron in a rotating orbit is constantly accelerating because the direction of the velocity is changing with time, even if the magnitude is not. Therefore, it is not enough to transform the laboratory frame to the rest frame using the electron's instantaneous velocity. An observer in the electron's rest frame finds that an additional rotation is required to align her/his coordinate axes with the ones obtained by Lorentz transforming the laboratory frame. If this fact is taken into account, it introduces an additional factor of 2 in the denominator of the expression above. Thomas' derivation

is exceedingly complex and omitted here, but the interested reader is referred to ref. [1].

The correct expression for the magnetic flux density is therefore

$$\vec{\mathcal{B}} = \frac{\vec{\mathcal{E}} \times \vec{v}}{2c^2\sqrt{1 - v^2/c^2}}. \tag{6.2}$$

If we write the magnetic moment of the self rotating electron as $\vec{\mu}_e$, then the energy of its interaction with $\vec{\mathcal{B}}$ is (recall the discussion in Chapter 2):

$$E_{rel} = -\vec{\mu_e} \cdot \vec{\mathcal{B}}. \tag{6.3}$$

We can call this interaction "spin-orbit interaction" since $\vec{\mu_e}$ arises from the "spin" and $\vec{\mathcal{B}}$ arises from the orbital motion.

The ratio of the magnetic moment μ_e (in units of the Bohr magneton μ_B) to the angular momentum of self rotation (in units of $\hbar$) is defined as the gyromagnetic ratio (or the Landé g-factor) g_0:

$$\frac{magnetic\ moment}{angular\ momentum} = -g_0\ . \tag{6.4}$$

Therefore,

$$\vec{\mu_e} = -g_0\mu_B\vec{s}, \tag{6.5}$$

since the angular momentum of self rotation in units of $\hbar$ is equal to $\vec{s}$.

As a result, Equation (6.3) can be re-written as

$$E_{rel} = g_0\mu_B\vec{s} \cdot \vec{\mathcal{B}}. \tag{6.6}$$

Using Equation (6.2), the above equation can be re-cast as

$$E_{rel} = g_0\mu_B\frac{\vec{E} \times \vec{v}}{2c^2\sqrt{1 - v^2/c^2}} \cdot \vec{s} = \frac{g_0}{2}\frac{e\hbar}{2m}\frac{\vec{E} \times \vec{v}}{2c^2\sqrt{1 - v^2/c^2}} \cdot \vec{\sigma}, \tag{6.7}$$

where we have used the relation $\vec{s} = (1/2)\vec{\sigma}$ and the relation $\mu_B = e\hbar/(2m)$.

For a non-relativistic electron, i.e., an electron orbiting around a nucleus with an orbital velocity v_{orbit} much less than the speed of light in vacuum ($v_{orbit}^2/c^2 << 1$), the non-relativistic version of Equation (6.2) can be derived in another way. An observer sitting on the electron and moving with it will think that the electron is at rest and the nucleus is revolving around it with a velocity -$\vec{v}_{orbit}$. To the observer, the radius of the orbit will be -$\vec{r}$ and the nuclear charge will be $+Ze$. Consequently, according to Biot Savart's law, the magnetic flux density at the position of the electron will be given by

$$\vec{B_0} = Ze\frac{\vec{r} \times \vec{v_{orbit}}}{4\pi\epsilon_0 c^2 r^3}, \tag{6.8}$$

where ϵ_0 is the permittivity of free space.

Since the electric field $\mathcal{E}$ seen by the orbiting electron is the Coulomb field $Ze\vec{r}/(4\pi\epsilon_0 r^3)$, the above equation can be written as

$$B_0 = \frac{\vec{\mathcal{E}} \times \vec{v}}{c^2} \ , \tag{6.9}$$

which is the same as Equation (6.2) except for the Thomas factor of 2 in the denominator and the relativistic factor $\sqrt{1 - v^2/c^2}$. Therefore, for a non-relativistic electron, the magnetic field appearing at the orbiting electron can be derived from Biot Savart's law without invoking Lorentz transformation. Of course, we will have to then add the Thomas correction.

Note that Equation (6.8) can be written as

$$\vec{B_0} = Ze\frac{\vec{K}}{4\pi\epsilon_0 mc^2 r^3}, \tag{6.10}$$

where $\vec{K}$ is the orbital angular momentum ($\vec{K} = m\vec{r} \times \vec{v}_{orbital}$) and m is the electron's mass. Since $\vec{K}$ is quantized in units of $\hbar$ ($\vec{K} = \vec{l}\hbar$), we can re-write the above equation as

$$\vec{B_0} = \frac{Ze\vec{l}\hbar}{4\pi\epsilon_0 mc^2 r^3}. \tag{6.11}$$

Substituting the above expression for $\vec{\mathcal{B}}$ in Equation (6.6), we get

$$E'_{rel} = g_0\mu_B\hbar\frac{Ze}{4\pi\epsilon_0 mc^2 r^3}\vec{l}\cdot\vec{s}. \tag{6.12}$$

After we add the Thomas correction, the above equation becomes

$$E'_{rel} = g_0\mu_B\hbar\frac{Ze}{8\pi\epsilon_0 mc^2 r^3}\vec{l}\cdot\vec{s}. \tag{6.13}$$

The above energy depends on the scalar product $\vec{l}\cdot\vec{s}$ where $\vec{l}$ is the orbital quantum number and $\vec{s}$ is the spin quantum number. Hence the name "spin-orbit interaction". The general form of this energy is still given by Equation (6.7).

Since Dirac had shown that $g_0 = 2$ for a free electron, the spin-orbit interaction Hamiltonian will be (see Equation (6.7)):

$$H_{so} = -\frac{e\hbar}{4m^2c^2\sqrt{1 - v^2/c^2}}\left(\vec{\nabla}V \times \vec{p}\right)\cdot\vec{\sigma} \approx -\frac{e\hbar}{4m^2c^2}\left(\vec{\nabla}V \times \vec{p}\right)\cdot\vec{\sigma}, \tag{6.14}$$

where the electric field is related to the electric potential V as $\vec{\mathcal{E}} = -\vec{\nabla}V$ and the velocity operator is $\vec{p}/m$ where $\vec{p} = -i\hbar\vec{\nabla}$ is the momentum operator.

6.1 Spin orbit interaction in a solid

In a solid, a quasi-free electron in the conduction band does not experience the strong nuclear attraction as in an atom. However, it may still see an electric field (or potential gradient) due to internal effects. For example, if there is a strong electric field due to an internal potential gradient (such as that associated with the conduction band discontinuity in a heterostructure), or due to an externally applied electric field, then this field will induce a spin-orbit interaction of the type given by Equation (6.14). This problem was examined by E. I. Rashba in 1960. The associated spin-orbit interaction bears his name (Rashba interaction) and is also sometimes referred to as spin-orbit interaction due to structural inversion asymmetry (SIA) since an external or internal electric field is responsible for this interaction (because it gives rise to an inversion asymmetry in the conduction band profile). Another type of internal electric field may arise due to crystallographic inversion asymmetry in a crystal. This problem was examined by G. Dresselhaus in 1954 and the associated spin-orbit interaction bears his name (Dresselhaus interaction). Here, we will examine these two effects.

6.1.1 Rashba interaction

Since any three vectors $\vec{A}$, $\vec{B}$ and $\vec{C}$ obey the relation

$$\left(\vec{A}\times\vec{B}\right)\cdot\vec{C} = -\vec{A}\cdot\left(\vec{C}\times\vec{B}\right)\ , \tag{6.15}$$

we can write the spin-orbit interaction Hamiltonian in a solid as

$$H_{so}^{correct} = \frac{ge\hbar}{8(m^*)^2c^2}\vec{\nabla}V\cdot(\vec{\sigma}\times\vec{p})\,, \tag{6.16}$$

where we have replaced the free electron mass m with the effective mass m^* since we are in a crystal, and omitted the relativistic factor $\sqrt{1-v^2/c^2}$ since we are dealing with electrons that move with speeds much smaller than the speed of light in vacuum. We have also reintroduced the g-factor of the material.

The Rashba interaction is obtained from Equation (6.16) in a straightforward way:

$$H_R = -\frac{ge\hbar}{8(m^*(\vec{r}))^2c^2}\vec{\mathcal{E}}(\vec{r})\cdot(\vec{\sigma}\times\vec{p}) = \vec{\eta}_R(\vec{r})\cdot(\vec{\sigma}\times\vec{p})\,, \tag{6.17}$$

where

$$\vec{\eta}_R(\vec{r}) = -\frac{ge\hbar}{8(m^*(\vec{r}))^2c^2}\vec{\mathcal{E}}(\vec{r}). \tag{6.18}$$

The above equation is of course simplistic since we have neglected band structure effects in a solid. If we account for that, then [3]

$$\vec{\eta}_R(\vec{r}) = -\frac{e\hbar}{m^*(\vec{r})}\frac{\pi\Delta_s(2E_g+\Delta_s)}{E_g(E_g+\Delta_s)(3E_g+2\Delta_s)}\vec{\mathcal{E}}(\vec{r}), \tag{6.19}$$

where E_g is the band gap of the semiconductor and Δ_s is the spin-orbit splitting in the valence band. Here, we have also made allowance for the fact that the effective mass of an electron in a heterostructure can vary spatially. There are some subtle issues regarding the spatial variation of $\vec{\eta}_R$, which we will not discuss. They are outside the scope of this textbook.

In the presence of a magnetic field, the momentum operator $\vec{p}$ must be replaced by $\vec{p}+e\vec{A}$, where $\vec{A}$ is the vector potential associated with the magnetic field. In that case, the general expression for the Hamiltonian describing Rashba interaction in a magnetic field is

$$H_R = \vec{\eta}_R(\vec{r})\cdot\left[\vec{\sigma}\times(\vec{p}+e\vec{A})\right]. \tag{6.20}$$

Another popular form for the Rashba interaction is

$$H_R = \frac{a_{46}}{\hbar}\vec{\mathcal{E}}(\vec{r})\cdot\left[\vec{\sigma}\times(\vec{p}+e\vec{A})\right], \tag{6.21}$$

where a_{46} is a material constant. Obviously, $\frac{a_{46}}{\hbar}\vec{\mathcal{E}}(\vec{r}) = \vec{\eta}_R(\vec{r})$. Therefore,

$$a_{46} = \frac{\pi e\hbar^2}{m^*}\frac{\Delta_s(2E_g+\Delta_s)}{E_g(E_g+\Delta_s)(3E_g+2\Delta_s)}, \tag{6.22}$$

where E_g is the bandgap and Δ_s is the spin-orbit splitting energy in the valence band. In GaAs, the value of a_{46} has been reported to be of the order of 10^{-38} C-m^2 [4].

Using this type of spin-orbit interaction, Rashba and Bychkov [2] were able to explain many features observed in heterostructures such as the combined resonances (electric dipole resonance + cyclotron resonance) observed in two-dimensional electron gases formed at AlGaAs/GaAs hetero-interfaces. These experiments were reported in the early 1980s by Stein [5] and Stormer [6] and suggested the existence of a spin-splitting due to the Rashba spin-orbit interaction. Additionally, the observation of beating patterns in Shubnikov-deHaas oscillations in two-dimensional electron gases formed at InAs/InAlAs hetero-interfaces was explained by the spin-splitting of the lowest subband due to the Rashba interaction [7]. The Rashba interaction plays an extremely important role in spintronics since this interaction can be *tuned* by an external electric field $\vec{\mathcal{E}}$ (see Equation (6.18)). As such, it is the basis of many spintronic devices - both classical such as the Spin Field Effect Transistor [8], and quantum mechanical such as universal spin-based quantum gate [9].

6.1.2 Dresselhaus interaction

The Pauli Equation describing a single electron in a crystal lattice in the absence of any external magnetic field, but in the presence of spin-orbit interaction is

$$\left[\frac{|\vec{p}|^2}{2m_0} + V_{lattice} - \frac{e\hbar}{4m_0^2c^2}\left(\vec{\nabla}V \times \vec{p}\right)\cdot\vec{\sigma}\right][\psi] = E[\psi]\ , \tag{6.23}$$

where $V_{lattice}$ is the (spatially periodic) electrostatic potential energy due to the ionized atoms in the crystal.

Since $V_{lattice}$ is periodic in space, the 2-component wavefunction should have the Bloch form

$$[\psi] = e^{i\vec{k}\cdot\vec{r}}\left[u_{\vec{k}}\left(\vec{r}\right)\right] = e^{i\vec{k}\cdot\vec{r}}\begin{bmatrix} u_{\vec{k}}^{\uparrow}\left(\vec{r}\right) \\ u_{\vec{k}}^{\downarrow}\left(\vec{r}\right)\end{bmatrix}\ . \tag{6.24}$$

Substituting Equation (6.24) in Equation (6.23), we get

$$\begin{aligned}\left[\frac{|\vec{p}|^2}{2m_0} + V_{lattice} - \frac{e\hbar}{4m_0^2c^2}\left(\vec{\nabla}V \times \vec{p}\right)\cdot\vec{\sigma}\right]\left[u_{\vec{k}}\left(\vec{r}\right)\right] \\ +\hbar\vec{k}\cdot\left[\frac{|\vec{p}|}{m_0} - \frac{e\hbar}{4m_0^2c^2}\left(\vec{\sigma}\times\vec{\nabla}V\right)\right]\left[u_{\vec{k}}\left(\vec{r}\right)\right] = \left[E_{\vec{k}} - \frac{\hbar^2k^2}{2m_0}\right]\left[u_{\vec{k}}\left(\vec{r}\right)\right].\end{aligned} \tag{6.25}$$

The equation for $\vec{k}+\vec{K}$ (where $\vec{K}$ is reciprocal lattice vector) is

$$\begin{aligned}\left[\frac{|\vec{p}|^2}{2m_0} + V_{lattice} - \frac{e\hbar}{4m_0^2c^2}\left(\vec{\nabla}V \times \vec{p}\right)\cdot\vec{\sigma}\right]\left[u_{\vec{k}+\vec{K}}\left(\vec{r}\right)\right] \\ +\hbar\vec{k}\cdot\left[\frac{|\vec{p}|}{m_0} - \frac{e\hbar}{4m_0^2c^2}\left(\vec{\sigma}\times\vec{\nabla}V\right)\right]\left[u_{\vec{k}+\vec{K}}\left(\vec{r}\right)\right] \\ +\hbar\vec{K}\cdot\left[\frac{|\vec{p}|}{m_0} - \frac{e\hbar}{4m_0^2c^2}\left(\vec{\sigma}\times\vec{\nabla}V\right)\right]\left[u_{\vec{k}+\vec{K}}\left(\vec{r}\right)\right] \\ = \left[E_{\vec{k}} - \frac{\hbar^2\left(\vec{k}+\vec{K}\right)^2}{2m_0}\right]\left[u_{\vec{k}+\vec{K}}\left(\vec{r}\right)\right].\end{aligned} \tag{6.26}$$

Treating the term $\hbar\vec{K}\cdot\left[\frac{\vec{p}}{m_0} - \frac{e\hbar}{4m_0^2c^2}\left(\vec{\sigma}\times\vec{\nabla}V\right)\right]$ as a perturbation and applying group theoretic results, Dresselhaus was able to derive the spin-orbit interaction Hamiltonian and spin-splitting energies in principal crystallographic directions (e.g., [100], [111], etc.) [10]. Dresselhaus' complete derivation is too complex and involved to reproduce here and the reader is referred to ref. [10] for more details.

For the [100] crystallographic direction, the Dresselhaus spin-orbit interaction Hamiltonian becomes

$$H_D = a_{42}\sigma \cdot \vec{\kappa} \ , \tag{6.27}$$

where

$$\begin{aligned}
\kappa_x &= \frac{1}{2\hbar^3}\left[p_x\left(p_y^2 - p_z^2\right) + \left(p_y^2 - p_z^2\right)p_x\right] \\
\kappa_y &= \frac{1}{2\hbar^3}\left[p_y\left(p_z^2 - p_x^2\right) + \left(p_z^2 - p_x^2\right)p_y\right] \\
\kappa_z &= \frac{1}{2\hbar^3}\left[p_z\left(p_x^2 - p_y^2\right) + \left(p_x^2 - p_y^2\right)p_z\right] \ ,
\end{aligned} \tag{6.28}$$

and a_{42} is a material constant. If there is a magnetic field with an associated vector potential $\vec{A}$, then as usual the momentum operator in the above equation should be transformed according to

$$\vec{p} \rightarrow \vec{p} + e\vec{A}. \tag{6.29}$$

6.2 Problems

- **Problem 6.1**

  Consider an electron orbiting a proton in a circular loop. Show using Biot Savart's law that the orbiting electron experiences a magnetic flux density of $\vec{B} = \frac{1}{2}\frac{\vec{E}\times\vec{v}}{c^2}$, where we have inserted the Thomas correction of a factor of 2 in the denominator.

  Solution

  If we sit on the electron, the proton will appear to rotate around us and create a circular loop of current which is given by

  $$I = \frac{e}{T} = \frac{ev}{2\pi R}, \tag{6.30}$$

  where v is the velocity of the proton moving on the circular orbit, R is the radius of the circular orbit and T is the period of rotation which is identical to the period of rotation of the electron around the heavy proton in the rest frame of the latter.

  According to Biot-Savart's law, the magnetic flux density at a distance x along a line passing through the center of the proton orbit and perpendicular to the plane of the orbit is given by

  $$B = \frac{\mu_0 I R^2}{2\sqrt{R^2 + x^2}^{3/2}}. \tag{6.31}$$

So, in the plane of the orbit, where $x = 0$, we have

$$B = \frac{\mu_0 I}{2R} = \frac{\mu_0 e}{2RT} = \frac{\mu_0 ev}{4\pi R^2}. \tag{6.32}$$

Furthermore, since $c^2 = \frac{1}{\mu_0 \epsilon_0}$, we get

$$B = \frac{Ev}{c^2}, \tag{6.33}$$

which, in the rest frame of the electron, can be written in a vectorial form as

$$\vec{B} = \frac{1}{2}\frac{\vec{E} \times \vec{v}}{c^2}\ , \tag{6.34}$$

where we have inserted the Thomas "factor of 2 correction" by hand.

- **Problem 6.2**

  Consider a two-dimensional electron gas in the x-z plane and a symmetry breaking electric field $\mathcal{E}_y$ in the y-direction, which induces Rashba spin-orbit interaction. Show that the Rashba Hamiltonian is

$$[H_R] = \frac{a_{46}}{\hbar}\mathcal{E}_y \left(p_x\sigma_z - p_z\sigma_x\right). \tag{6.35}$$

  Assume now that there is a magnetic flux density B in the y-direction as well. In that case, use the Landau gauge $\vec{A} = (Bz, 0, 0)$, to show that the Rashba Hamiltonian is

$$\begin{aligned}[H_R] &= \frac{a_{46}}{\hbar}\mathcal{E}_y \left[(p_x + eBz)\sigma_z - p_z\sigma_x\right] \\ &= \frac{a_{46}}{\hbar}\mathcal{E}_y \begin{bmatrix} p_x + eBz & -p_z \\ -p_z & -p_x - eBz \end{bmatrix}. \end{aligned} \tag{6.36}$$

- **Problem 6.3**

  Derive Equations (6.25) and (6.26). Hint: $\vec{p} = -i\hbar\nabla$ and $\nabla\left[e^{i\vec{k}\cdot\vec{r}}u_{\vec{k}}(\vec{r})\right] = i\vec{k}\left[e^{i\vec{k}\cdot\vec{r}}u_{\vec{k}}(\vec{r})\right] + e^{i\vec{k}\cdot\vec{r}}\nabla u_{\vec{k}}(\vec{r})$.

6.3 References

[1] L. H. Thomas, "The motion of the spinning electron", Nature, **117**, 514 (1926). For a slightly easier derivation, see H. Kroemer, "The Thomas precession factor in spin orbit interaction", Am. J. Phys., **72**, 51 (2004) and J. A. Rhodes and M. D. Semon, "Relativistic velocity space, Wigner rotation, and Thomas precession", Am. J. Phys., **72**, 943 (2004).

[2] Yu A. Bychkov and E. I. Rashba, "Oscillatory effects and the magnetic susceptibility of carriers in inversion layers", J. Phys. C, **17**, 6039 (1984).

[3] F. G. Pikus and G. E. Pikus, "Conduction band spin splitting and negative magnetoresistance in A_3B_5 heterostructures", Phys. Rev. B, **51**, 16928 (1995).

[4] A. Bournel, P. Dollfus, S. Galdin, F. X. Musalem and P. Hesto, "Modelling of gate induced spin precession in a striped channel high electron mobility transistor", Solid St. Commun., **104**, 85 (1997).

[5] D. Stein, K. v. Klitzing and G. Weimann, "Electron spin resonance on GaAs-$Al_xGa_{1-x}As$ heterostructures", Phys. Rev. Lett., **51**, 130 (1983).

[6] H. L. Störmer, Z. Schlesinger, A. Chang, D. C. Tsui, A. C. Gossard and W. Weigmann, "Energy structure and quantized Hall effect of two dimensional holes", Phys. Rev. Lett., **51**, 126 (1983).

[7] B. Das, D. C. Miller, S. Datta, R. Reifenberger, W. P. Hong, P. K. Bhattacharyya, J. Singh and M. Jaffe, "Zero field spin splitting in a two-dimensional electron gas", Phys. Rev. B, **39**, 1411 (1989).

[8] S. Datta and B. Das, "Electronic analog of the electro-optic modulator", Appl. Phys. Lett., **56**, 665 (1990).

[9] S. Bandyopadhyay, "Self assembled nanoelectronic quantum computer based on the Rashba effect in quantum dots", Phys. Rev. B, **61**, 13813 (2000).

[10] G. Dresselhaus, "Spin orbit coupling effects in zinc blende structures", Phys. Rev., **100**, 580 (1955).

7

Magneto-Electric Subbands in Quantum Confined Structures in the Presence of Spin-Orbit Interaction

When an electron is confined in a structure of restricted dimensionality (such as a quantum well or a quantum wire) and is subjected to a magnetic field, the electron experiences both *electrostatic confinement* due to structure and *magnetostatic confinement* due to the magnetic field. As a result, the allowed electronic states are *hybrid magneto-electric states* which form *magneto-electric subbands*. In this chapter, we will derive the dispersion relations (energy versus wavevector) of electrons in magneto-electric subbands in the presence of *spin-orbit interaction*. The two types of spin-orbit interactions that we will consider are the common ones in a solid - the Rashba and the Dresselhaus interactions.

7.1 Dispersion relations of spin resolved magneto-electric subbands and eigenspinors in a two-dimensional electron gas in the presence of spin-orbit interaction

Consider a two-dimensional electron gas (2-DEG) such as the one encountered at the interface of a heterostructure or a quantum well. Figure 7.1 shows such a system. We will assume that there is a symmetry breaking electric field $\mathcal{E}_y$ along the y-axis that induces a Rashba spin-orbit interaction. This electric field could come about because of the conduction band discontinuity at the hetero-interface between two different materials, or could be applied from outside by attaching a gate electrode and connecting it to the floating terminal of a grounded voltage source (see the discussion of the Spin Field Effect Transistor in Chapter 12). In addition to the Rashba interaction, there is also the Dresselhaus interaction accruing from bulk (or crystallographic) inversion asymmetry. Finally, there is also an external magnetic field inducing the Zeeman interaction that we encountered in Chapter 2. Our intention is to derive the dispersion relation (relation between electron kinetic energy and wavevector) in this two-dimensional system. We will also derive the

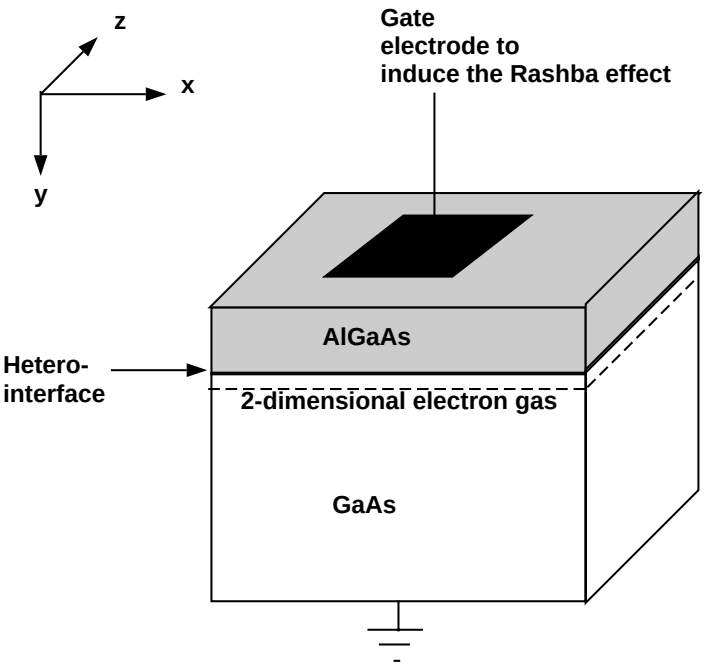


FIGURE 7.1
A heterostructure containing a two-dimensional electron gas.

eigenspinors (or spin eigenstates) at any given wavevector in a given subband. Since the boundaries of the quantum confined structure restrict motion in the direction perpendicular to the boundary, and the external magnetic field also constrains the electron's motion by forcing it into closed Landau orbits, the electron experiences both electrostatic and magnetostatic confinement. This discretizes the allowed energies, breaking it up into discrete subbands. They are called *magneto-electric subbands* since they are due to the combined effect of electrostatic and magnetostatic confinements.

In this Chapter, we will find the energy dispersion relations (electron energy E as a function of electron wavevector k) in the magneto-electric subbands when Rashba and Dresselhaus spin-orbit interactions are present. We will find that generally the eigenspinors (and therefore the spin orientation) will be wavevector dependent. We will find the eigenspinors as a function of the wavevector in different subbands.

Because of the wavevector dependence of spin orientation, whenever an electron changes momentum (or wavevector) owing to scattering, its spin orientation also changes. Consequently, momentum relaxation simultaneously causes spin relaxation. This is the basis of a spin relaxation mechanism discussed in ref. [1] which arises in the presence of both spin-orbit interaction and an external magnetic field. It is similar to the Elliott-Yafet mechanism of spin relaxation discussed in Chapter 8.

Consider the quantum well containing a two-dimensional electron gas (2-DEG) as shown in Fig. 7.1. Assume that both Rashba and Dresselhaus spin-orbit interactions are present. Additionally, there is an external magnetic field lying either in the plane of the 2-DEG, or directed perpendicular to it.

The Rashba spin-orbit interaction has the Hamiltonian

$$H_R = \vec{\eta}_R(\mathcal{E}) \cdot \left[\vec{\sigma} \times (\vec{p} + e\vec{A})\right], \tag{7.1}$$

where $\vec{A}$ is the vector potential associated with the magnetic field, $\vec{\eta}_R(\mathcal{E})$ is a vector whose magnitude is the strength of the Rashba interaction. This strength depends on the magnitude of the symmetry breaking electric field $\vec{\mathcal{E}}$ and the direction of $\vec{\eta}_R(\mathcal{E})$ is along $\vec{\mathcal{E}}$. Finally, $\vec{\sigma}$ is the Pauli spin matrix and $\vec{p}$ is the momentum operator.

The Hamiltonian for the Dresselhaus interaction can be complicated, but for a 2-DEG in a principal crystallographic plane (e.g., [100] plane), it usually has the form:

$$H_D = \nu_D \vec{\sigma} \cdot \vec{\kappa}, \tag{7.2}$$

where ν is the strength of the Dresselhaus interaction ($\nu_D = a_{42}$) and

$$\begin{aligned}
\kappa_x &= \frac{1}{2\hbar^3}[(p_x + eA_x)\left\{(p_y + eA_y)^2 - (p_z + eA_z)^2\right\} \\
&\quad + \left\{(p_y + eA_y)^2 - (p_z + eA_z)^2\right\}(p_x + eA_x)] \\
\kappa_y &= \frac{1}{2\hbar^3}[(p_y + eA_y)\left\{(p_z + eA_z)^2 - (p_x + eA_x)^2\right\} \\
&\quad + \left\{(p_z + eA_z)^2 - (p_x + eA_x)^2\right\}(p_y + eA_y)] \\
\kappa_z &= \frac{1}{2\hbar^3}[(p_z + eA_z)\left\{(p_x + eA_x)^2 - (p_y + eA_y)^2\right\} \\
&\quad + \left\{(p_x + eA_x)^2 - (p_y + eA_y)^2\right\}(p_z + eA_z)].
\end{aligned} \tag{7.3}$$

Note that by using the symmetric combination for the κ-s, we ensure that the Hamiltonian H_D is Hermitean. Note also that the operators $(p_i + eA_i)[(p_j + eA_j)^2 - (p_k + eA_k)^2]$ and $[(p_j + eA_j)^2 - (p_k + eA_k)^2](p_i + eA_i)$ do not necessarily commute; hence, it is imperative to use the symmetric combination to generate a Hermitean operator.

The general Hamiltonian describing the 2-DEG system in a magnetic field

of flux density $\vec{B}$ is

$$H_{2-DEG} = \frac{\left|\vec{p} + e\vec{A}\right|^2}{2m^*} + V(y), + H_Z + H_R + H_D \tag{7.4}$$

where the electrons are confined along the y-direction by the electrostatic confining potential $V(y)$, H_Z is the Zeeman interaction Hamiltonian and $\vec{A}$ is the vector potential associated with the magnetic field such that

$$\vec{\nabla} \times \vec{A} = \vec{B}. \tag{7.5}$$

7.1.1 Magnetic field in the plane of the 2-DEG

We are now ready to derive the energy dispersion $(E - k)$ relation for the 2-DEG. First consider the situation when there is a magnetic field in the plane of the 2-DEG. In this case, the magnetic flux density is given by $\vec{B} = B_x\hat{x} + B_z\hat{z}$ since the 2-DEG is in the x-z plane.

We choose the gauge $\vec{A} = -B_z y\hat{x} + B_x y\hat{z}$. If the 2-DEG has vanishing thickness, then the spatial average $< p_y^2 > >> < p_x^2 >, < p_z^2 >$, and the effective mass Hamiltonian describing the 2-DEG can be written in the form

$$\begin{aligned} H_{2-DEG}^{||} \approx & \frac{1}{2m^*}\left[(p_x - eB_z y)^2 + p_y^2 + (p_z + eB_x y)^2\right] + V(y) \\ & -\frac{g}{2}\mu_B\left[B_x\sigma_x + B_z\sigma_z\right] - \frac{\eta}{\hbar}\left[(p_x - eB_z y)\,\sigma_z - (p_z + eB_x y)\,\sigma_x\right] \\ & \frac{\nu_D}{\hbar}\frac{p_y^2}{\hbar^2}\left[(p_x - eB_z y)\,\sigma_x - (p_z + eB_x y)\,\sigma_z\right], \end{aligned} \tag{7.6}$$

where $V(y)$ is the confining potential in the y-direction (including the effect of the y-directed electric field causing the Rashba interaction) and η = - $\hbar|\vec{\eta}_R(\mathcal{E})|$. The quantity η is related to the material constant a_{46} as $\eta = a_{46}\mathcal{E}$ where $\mathcal{E}$ is the symmetry-breaking electric field inducing the Rashba effect. In InAs heterostructures with built-in electric fields of the order of 100 kV/cm, η is of the order of 10^{-12} - 10^{-11}eV-m.

The terms in the second and third line of the last equation are spin-dependent terms since they involve the Pauli spin matrix. Note that the first of these spin-dependent terms is the Zeeman interaction, the second is the Rashba interaction and the last is the Dresselhaus interaction.

To find the energy dispersion relations in this 2-DEG and the 2-component wavefunction for the spin eigenstates, we have to solve the non-relativistic *Pauli Equation* (recall Chapter 2 material):

$$\left[H_{2-DEG}^{||}\right]\left[\psi_{2-DEG}^{||}(x,y,z)\right] = E\left[\psi_{2-DEG}^{||}(x,y,z)\right], \tag{7.7}$$

where $\left[H_{2-DEG}^{||}\right]$ is a 2×2 matrix and $\left[\psi_{2-DEG}^{||}(x,y,z)\right]$ is the 2-component wavefunction which is a 2×1 column vector. The superscript $||$ reminds the reader that there is a magnetic field parallel to the plane of the 2-DEG.

Since the Hamiltonian is invariant in x and z, the wavevectors k_x and k_z are good quantum numbers and we can write the 2-component wavefunction as

$$\left[\psi^{||}_{2-DEG}(x,y,z)\right] = e^{ik_x x}e^{ik_z z}[\lambda(y)], \tag{7.8}$$

where $[\lambda(y)]$ is a 2×1 column vector.

The Pauli Equation yields

$$\begin{aligned}
\{\frac{\hbar^2 k_x^2}{2m^*} + \frac{\hbar^2 k_z^2}{2m^*} - \frac{\hbar^2}{2m^*}\frac{\partial^2}{\partial y^2} + V(y) \\
-eB_z y\frac{\hbar k_x}{m^*} + eB_x y\frac{\hbar k_z}{m^*} + e^2y^2(B_x^2 + B_z^2)/(2m^*) \\
-\frac{g}{2}\mu_B\left[B_x\sigma_x + B_z\sigma_z\right] \\
-\frac{\eta}{\hbar}\left[(\hbar k_x - eB_z y)\,\sigma_z - (\hbar k_z + eB_x y)\,\sigma_x\right] \\
+\frac{\nu_D}{\hbar}\frac{p_y^2}{\hbar^2}\left[(\hbar k_x - eB_z y)\,\sigma_x - (\hbar k_z + eB_x y)\,\sigma_z\right]\}[\lambda(y)] = E[\lambda(y)],
\end{aligned} \tag{7.9}$$

and the boundary conditions are

$$[\lambda(y)](y = d) = [\lambda(y)](y = -d) = [0], \tag{7.10}$$

where the width of the well in the y-direction is assumed to be $2d$, and $[0]$ is the null vector. Here we have assumed a rectangular potential well of infinite barrier for the sake of mathematical simplicity. For a triangular potential well that is usually found at the interface of two semiconductors of different bandgaps as in Fig. 7.1, the boundary conditions will be, of course, different. If the barriers are finite, then the problem is solved by enforcing the continuity of the wavefunction and its first derivative across the interface between the well and the barrier materials.

In order to find the dispersion relation, we will have to find the values of the wavevector k_x that satisfy the above two equations for a given value of k_z and E. We then repeat this for various values of E and the same value of k_z to find the dispersion relation (E versus k_x) for a fixed value of k_z. This procedure can be continued to find E versus k_x for different values of k_z, which will finally yield E versus $\vec{k}_t = k_x\hat{x} + k_z\hat{z}$.

Unfortunately, this is not straightforward since Equation (7.9) is not an eigenequation in k_x because it is non-linear in k_x. We therefore have to convert Equation (7.9) to an eigenequation in k_x using the following procedure.

Let

$$[\zeta(y)] = k_x[\lambda(y)]. \tag{7.11}$$

Equation (7.9) can be written as

$$\{\mathbf{B}k_x + \mathbf{C}\}\,[\lambda(y)] = \mathbf{A}k_x^2[\lambda(y)], \tag{7.12}$$

where the 2×2 matrices $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ are given by

$$\mathbf{A} = -\frac{\hbar^2}{2m^*}\mathbf{I}, \tag{7.13}$$

$$\mathbf{B} = -eB_z y\frac{\hbar}{m^*}\mathbf{I} - \eta[\sigma_{\mathbf{z}}] - \nu[\sigma_{\mathbf{x}}], \tag{7.14}$$

$$\begin{aligned}\mathbf{C} = &\frac{\hbar^2 k_z^2}{2m^*}\mathbf{I} - \frac{\hbar^2}{2m^*}\frac{\partial^2}{\partial y^2}\mathbf{I} + eB_x y\frac{\hbar k_z}{m^*}\mathbf{I} + V(y)\mathbf{I} \\ &+\frac{e^2y^2}{2m^*}\left(B_x^2 + B_z^2\right)\mathbf{I} - \frac{g}{2}\mu_B\left\{B_x[\sigma_{\mathbf{x}}] + B_z[\sigma_{\mathbf{z}}]\right\} - E\mathbf{I} \\ &+\frac{eB_z y}{\hbar}\left[\eta[\sigma_{\mathbf{z}}] + \nu_D\frac{\partial^2}{\partial y^2}[\sigma_{\mathbf{x}}]\right] \\ &+\left(k_z + \frac{eB_x y}{\hbar}\right)\left[\eta[\sigma_{\mathbf{x}}] + \nu_D\frac{\partial^2}{\partial y^2}[\sigma_{\mathbf{z}}]\right].\end{aligned} \tag{7.15}$$

Using the transformation in Equation (7.11), Equation (7.12) can be recast as an eigenequation in k_x

$$\begin{bmatrix} \mathbf{0} & \mathbf{I} \\ \{\mathbf{A}^{-1}\mathbf{C}\} & \{\mathbf{A}^{-1}\mathbf{B}\} \end{bmatrix}\begin{bmatrix}\lambda(y) \\ \zeta(y)\end{bmatrix} = k_x\begin{bmatrix}\lambda(y) \\ \zeta(y)\end{bmatrix}. \tag{7.16}$$

To find the eigenvalues k_x for a given E and k_z, we will have to use a numerical solution of Equation (7.16) subject to the boundary condition in Equation (7.10). If we actually carried this out, we will not only find the dispersion relation E_n versus k_x (in the n-th magneto-electric subband) for any given k_z, but we will also find the 2-component wavefunction $[\lambda(y)] = [\lambda_{+,n,k_x,k_z}(y), \lambda_{-,n,k_x,k_z}(y)]^\dagger$ in the n-th magneto-electric subband at wavevectors k_x and k_z. It is entirely possible that $\lambda_{+,n,k_x,k_z}(y) \neq \lambda_{-,n,k_x,k_z}(y)$, i.e., the spatial parts of the wavefunctions of the two spin eigenstates at given wavevectors k_x, k_z are *different*. In other words, the spatial wavefunction is spin-dependent.

While the above procedure for finding the dispersion relation is exact, it is also painstaking, cumbersome and does not yield an analytical result. In order to obtain an analytical result, we will make the following approximation.

First, we will assume that

$$[\lambda(y)] = \begin{bmatrix}\lambda_1(y) \\ \lambda_2(y)\end{bmatrix} = \lambda_0(y)\begin{bmatrix}\alpha \\ \beta\end{bmatrix}. \tag{7.17}$$

We call this the "variable separation" approximation, since we have written the eigenspinor as the *product* of a space dependent part $\lambda_0(y)$ and a spin-dependent part $[\alpha\ \beta]^\dagger$. The latter does not depend on the y-coordinate, which means that we are assuming (heuristically) that the spin orientation of an electron does not depend on the y-coordinate, or conversely, the spatial part of

the wavefunction is not spin-dependent. A consequence of this approximation is that we can miss some interesting and important physics, such as "spin texturing", whereby the spin orientation (and therefore the spin density) can vary along the y-direction [2, 3]. The second approximation that we make is that the width of the 2-DEG is so narrow that the subbands are well separated in energy. As a result, we can ignore subband mixing. With these two approximations, we are able to obtain an analytical result.

Note that with the variable separation approximation, the 2-component wavefunction can be written as

$$[\psi^{||}_{2-DEG}] = e^{ik_x x} e^{ik_z z} \lambda_0(y) \begin{bmatrix} \alpha \\ \beta \end{bmatrix}. \tag{7.18}$$

Let $< y > = \int_0^\infty \lambda_0^*(y) y \lambda_0(y) dy = y_0$, and without loss of generality, we can assume that $y_0 = 0$. In that case, the spatial average of the above Hamiltonian (spatial average is obtained by averaging over the x, y and z-coordinates) is

$$\begin{aligned} E_n &\approx \left\langle H^{||}_{2-DEG} \right\rangle = \left\langle H^{||} \right\rangle = \left\langle \psi^{||}_{2-DEG} | H_{2-DEG} | \psi^{||}_{2-DEG} \right\rangle \\ &= \epsilon_n + \frac{\hbar^2 \left(k_x^2 + k_z^2\right)}{2m^*} - \frac{g}{2}\mu_B \left[B_x \sigma_x + B_z \sigma_z\right] \\ &\quad -\eta \left[k_x \sigma_z - k_z \sigma_x\right] - \nu \left[k_x \sigma_x - k_z \sigma_z\right], \end{aligned} \tag{7.19}$$

where $\nu = -(\nu_D/\hbar^2) < p_y^2 > = \nu_D < \partial^2/\partial y^2 >$ and

$$\left[-\frac{\hbar^2}{2m^*} \frac{\partial^2}{\partial y^2} + \frac{e^2 y^2 (B_x^2 + B_z^2)}{2m^*} \right] \lambda_0^n(y) = \epsilon_n \lambda_0^n(y). \tag{7.20}$$

Writing the above equation explicitly in terms of the Pauli spin matrices, we get that

$$\begin{aligned} E_n[\mathbf{I}] &= \frac{\hbar^2 \left(k_x^2 + k_z^2\right)}{2m^*}[\mathbf{I}] + \epsilon_n[\mathbf{I}] - \frac{g}{2}\mu_B \begin{bmatrix} B_z & B_x \\ B_x & -B_z \end{bmatrix} \\ &\quad -\eta \begin{bmatrix} k_x & 0 \\ 0 & -k_x \end{bmatrix} + \eta \begin{bmatrix} 0 & k_z \\ k_z & 0 \end{bmatrix} - \nu \begin{bmatrix} 0 & k_x \\ k_x & 0 \end{bmatrix} + \nu \begin{bmatrix} k_z & 0 \\ 0 & -k_z \end{bmatrix} \\ &= \begin{bmatrix} \overline{E_n} - (g/2)\mu_B B_z - \eta k_x + \nu k_z & -(g/2)\mu_B B_x + \eta k_z - \nu k_x \\ -(g/2)\mu_B B_x + \eta k_z - \nu k_x & \overline{E_n} + (g/2)\mu_B B_z + \eta k_x - \nu k_z \end{bmatrix}, \end{aligned} \tag{7.21}$$

where

$$\overline{E_n} = \frac{\hbar^2 \left(k_x^2 + k_z^2\right)}{2m^*} + \epsilon_n. \tag{7.22}$$

In order to find the spin-dependent energy eigenstates E_n (as a function of k_x and k_z) for the n-th magneto-electric subband, as well as the corresponding eigenspinors, we have to "diagonalize the Hamiltonian", meaning that we must

find the eigenvalues and eigenfunctions of the matrix on the right hand side of Equation (7.21). Since we have a 2×2 matrix, there are two eigenvalues and two corresponding eigenfunctions. The two eigenvalues are the spin-dependent eigenenergies, and the two eigenfunctions are the eigenspinors. It is easy to evaluate the eigenvalues and eigenfunctions of the 2×2 matrix in Equation (7.21). We find that the eigenvalues in the n-th subband are:

$$E_{\pm}^{||} = \frac{\hbar^2 \left(k_x^2 + k_z^2\right)}{2m^*} + \epsilon_n \pm \sqrt{\left(\frac{g}{2}\mu_B B_z + \eta k_x - \nu k_z\right)^2 + \left(\frac{g}{2}\mu_B B_x - \eta k_z + \nu k_x\right)^2}. \tag{7.23}$$

The above equation gives the E versus (k_x, k_z) relations for the two spin-split levels in any subband n. Again the superscript '||' reminds the reader that there is an external magnetic field parallel to the plane of the 2-DEG.

The corresponding eigenspinors are:

$$\psi_{+}^{||}(B_x, B_z, k_x, k_z) = \begin{bmatrix} -sin(\theta_k) \\ cos(\theta_k) \end{bmatrix} \tag{7.24}$$

and

$$\psi_{-}^{||}(B_x, B_z, k_x, k_z) = \begin{bmatrix} cos\theta_k \\ sin\theta_k \end{bmatrix}, \tag{7.25}$$

where

$$\theta_k = \frac{1}{2} arctan \left[\frac{(g/2)\mu_B B_z - \eta k_z + \nu k_x}{(g/2)\mu_B B_x + \eta k_x - \nu k_z}\right]. \tag{7.26}$$

Note that at any given wavevector state (k_x, k_z), the two eigenspinors are orthogonal which means that the corresponding spins are anti-parallel. The angle θ_k however depends on the wavevector components k_x and k_z and is independent of the subband index n. This means that: (1) in any given subband, spins in the two spin split levels are *not* anti-parallel in *different* wavevector states but are anti-parallel in the same wavevector state, and (2) in any given wavevector state (k_x, k_z), spins in all the lower spin-split levels, in all subbands, are mutually parallel, as are the spins in all the upper spin-split levels.

Because of the wavevector dependence of the eigenspinors, neither spin-split level whose energy dispersion relation is given by Equation (7.23) has a fixed spin quantization axis. The spin quantization axis, or the spin polarization in any level, changes with changing wavevector. In Fig. 7.2, we plot the energy-versus wavevector relations (E vs. k_x, k_z) given by Equation (7.23) for the lowest magneto-electric subband ($n = 1$). Each magneto-electric subband is spin-split into two levels. The lower one has inflections along lines $k_x = 0$ and $k_z = 0$, whereas the upper one is nearly a paraboloid.

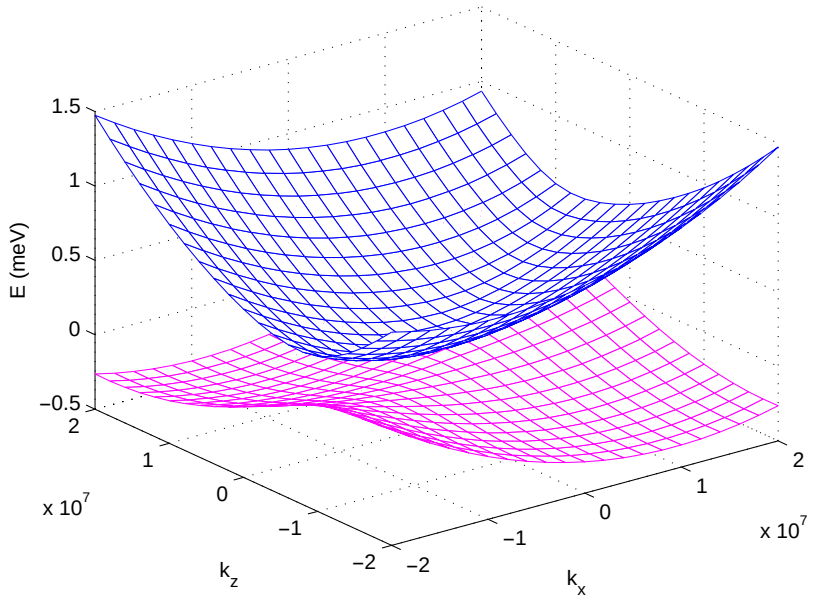


FIGURE 7.2
Energy dispersion relation of spin-split levels in the lowest subband in a two-dimensional electron gas subjected to a magnetic field in the plane of the gas. We assumed $m^* = 0.05m_0$, $g = 4$, $B = 1$ Tesla, $\eta = 10^{-11}$ eV-m, $\nu = 2\eta$. The wavevectors are in units of cm^{-1}.

7.1.1.1 Velocity versus wavevector relation

The velocity operator along the $\vec{q}$-direction is given by

$$v_q = \frac{\partial H^{||}_{2-DEG}}{\partial p_q} \ , \tag{7.27}$$

where $H^{||}_{2-DEG}$ is the Hamiltonian and p_q is the momentum along the $\vec{q}$-direction.

Using the Hamiltonian in Equation (7.6), we get that the expectation values of the velocity components along the x- and z-directions are

$$\begin{aligned} < v_x > &= \frac{< p_x >}{2m^*} - \frac{eB < y >}{m^*} - \frac{\eta}{\hbar} < \sigma_z > - \frac{\nu}{\hbar} < \sigma_x >, \\ < v_z > &= \frac{< p_z >}{2m^*} + \frac{eB < y >}{m^*} + \frac{\eta}{\hbar} < \sigma_x > + \frac{\nu}{\hbar} < \sigma_z > \ , \end{aligned} \tag{7.28}$$

where angular brakets $< ... >$ denote average over both the spatial and the spin parts of the wavefunctions. Using the eigenspinors in Equations (7.24) and (7.25) to calculate the expectation values of the velocity operators and dropping the angular brakets (they will henceforth be implied), we obtain that the velocities in the two spin-split levels in any wavevector state (k_x, k_z), in any subband, are different and given by

$$\begin{aligned} v_x^{\pm} &= \frac{\hbar k_x}{m^*} \pm \frac{\eta}{\hbar} cos(2\theta_k) \pm \frac{\nu}{\hbar} sin(2\theta_k), \\ v_z^{\pm} &= \frac{\hbar k_z}{m^*} \mp \frac{\nu}{\hbar} cos(2\theta_k) \mp \frac{\eta}{\hbar} sin(2\theta_k) \ , \end{aligned} \tag{7.29}$$

$$\tag{7.30}$$

where we have assumed, for convenience, that $< y > = 0$ since the origin of the y-coordinate is arbitrary.

Note that v_q is *not* proportional to k_q because of the additional terms that depend on the spin-orbit interaction strengths. Moreover, since θ_k depends on k_x and k_z, v_x depends on both k_x and k_z, as does v_z. This is expected since the dispersion relations are not parabolic.

7.1.1.2 Special case

Consider a situation when there is no external magnetic field and no Dresselhaus interaction. This is a situation that arises in the discussion of the Intrinsic Spin Hall Effect visited later. For this case, $tan(2\theta_k) = -k_z/k_x$, so that $sin(2\theta_k) = -k_z/k$ and $cos(2\theta_k) = k_x/k$, where $k = \sqrt{k_x^2 + k_z^2}$. Therefore, we obtain that

$$\begin{aligned} v_x^{\pm} &= \frac{\hbar k_x}{m^*} \pm \frac{\eta}{\hbar} \frac{k_x}{k} \\ v_z^{\pm} &= \frac{\hbar k_z}{m^*} \pm \frac{\eta}{\hbar} \frac{k_z}{k}. \end{aligned} \tag{7.31}$$

Therefore,

$$\left(v^{\pm}\right)^2 = \left(v_x^{\pm}\right)^2 + \left(v_z^{\pm}\right)^2 = \left(\frac{\hbar k}{m^*} \pm \frac{\eta}{\hbar}\right)^2$$
$$or \quad v^{\pm} = \pm\left(\frac{\hbar k}{m^*} \pm \frac{\eta}{\hbar}\right). \tag{7.32}$$

7.1.1.3 Another special case

There is one other special case of interest. Consider the situation if two conditions are fulfilled:

$$\begin{aligned} B_x &= B_z \\ \eta &= \nu \end{aligned} \tag{7.33}$$

meaning that the magnetic field subtends an angle of 45° with the x-axis, and the strengths of the Rashba and Dresselhaus interactions are equal. We can always choose our x and y axes orientation in a 2-DEG plane such that the magnetic field subtends an angle of 45° with either axes. All we need then is that the two spin-orbit interactions have equal strength. In that case

$$\theta_k = 22.5^\circ \tag{7.34}$$

and is independent of wavevector*.

In the above special case, each subband has a fixed quantization axis. We can find this spin quantization axis (spin polarization) easily. Let S_n be the component of spin polarization along the n-axis. Then

$$\begin{aligned} S_x^{\pm} &= (\hbar/2)[\psi_{\pm}]^T[\sigma_x][\psi_{\pm}] \\ S_y^{\pm} &= (\hbar/2)[\psi_{\pm}]^T[\sigma_y][\psi_{\pm}] \\ S_z^{\pm} &= (\hbar/2)[\psi_{\pm}]^T[\sigma_z][\psi_{\pm}] \end{aligned} \tag{7.35}$$

where $\psi_{\pm}$ is given by Equations (7.24) and (7.25). We thus obtain that $S_x^- = (\hbar/2)sin(45^\circ) = \hbar/(2\sqrt{2})$, $S_y^- = 0$ and $S_z^- = (\hbar/2)cos(45^\circ) = \hbar/(2\sqrt{2})$. Therefore, the spin polarization lies in the x-z plane and subtends an angle of 45° with either the x- or z-axis. In other words, it is directed along the magnetic field. The reader can easily verify that the spin polarization in the other spin-split level is *anti-parallel* to the spin polarization in the first level. Thus, the two spin polarizations are parallel and anti-parallel respectively to the magnetic field, as we expect from the bare Zeeman effect.

In this special case,

$$E_{\pm}^{||} = \frac{\hbar^2}{2m^*}\left(k_x \pm (m^*\eta)/\hbar^2\right)^2 + \frac{\hbar^2}{2m^*}\left(k_z \mp (m^*\eta)/\hbar^2\right)^2$$

*Note that θ_k becomes independent of wavevector also if $B_x = -B_z$ and $\eta = -\nu$.

$$
\begin{aligned}
&-\frac{m^*\eta^2}{\hbar^2} \pm (g/2)\mu_B B + \epsilon_n \\
&= \frac{\hbar^2}{2m^*}\left(k_x \pm (m^*\nu)/\hbar^2\right)^2 + \frac{\hbar^2}{2m^*}\left(k_z \mp (m^*\nu)/\hbar^2\right)^2 \\
&-\frac{m^*\nu^2}{\hbar^2} \pm (g/2)\mu_B B + \epsilon_n. \qquad (7.36)
\end{aligned}
$$

The velocity versus wavevector relations are:

$$
\begin{aligned}
v_x^{\pm} &= \frac{\hbar k_x}{m^*} \pm \sqrt{2}\frac{\eta}{\hbar} = \frac{\hbar k_x}{m^*} \pm \sqrt{2}\frac{\nu}{\hbar} \\
v_z^{\pm} &= \frac{\hbar k_z}{m^*} \mp \sqrt{2}\frac{\eta}{\hbar} = \frac{\hbar k_z}{m^*} \mp \sqrt{2}\frac{\nu}{\hbar}. \qquad (7.37)
\end{aligned}
$$

The velocity is linearly related to the wavevector. Moreover, the x-component of the velocities in the two spin-split levels (in any given subband) differ by a constant ($= 2\sqrt{2}\eta/\hbar = 2\sqrt{2}\nu/\hbar$) at the same value of the wavevector k_x. The same is true of the z-component of the velocity at the same value of the wavevector k_z.

7.1.2 Magnetic field perpendicular to the plane of the 2-DEG

When the magnetic field is perpendicular to the plane of the 2-DEG (in the y-direction), we can write the vector potential $\vec{A} = Bz\hat{x}$, where B is the flux density. In that case, the effective mass Hamiltonian is approximately given by

$$
\begin{aligned}
H_{2-DEG}^{\perp} = &\frac{(p_x + eBz)^2 + p_y^2 + p_z^2}{2m^*} + V(y) - \frac{g}{2}\mu_B B\sigma_y \\
&-\frac{\eta}{\hbar}\left[(p_x + eBz)\,\sigma_z - p_z\sigma_x\right] \\
&+\frac{\nu_D}{\hbar^3}p_y^2\left[(p_x + eBz)\,\sigma_x - p_z\sigma_z\right]. \qquad (7.38)
\end{aligned}
$$

Since the Hamiltonian is invariant in the coordinate x, the wavevector k_x is a good quantum number and we can write the spatial part of the wavefunction as

$$
\psi_{2-DEG}^{\perp} = e^{ik_x x}\phi(y)[\zeta(z)], \qquad (7.39)
$$

where $\phi(y)$ obeys the equation

$$
\left[\frac{p_y^2}{2m^*} + V(y)\right]\phi(y) = \epsilon\phi(y) \qquad (7.40)
$$

subject to the boundary condition $\phi(y = d) = \phi(y = -d) = 0$, and the 2-component wavefunction $[\zeta(z)]$ obeys the equation

$$
\{\frac{p_z^2}{2m^*} + \frac{m^*}{2}\omega_c^2\left(z + z_0\right)^2 - \frac{g}{2}\mu_B B\sigma_y
$$

$$-\frac{\eta}{\hbar}\left[m^*\omega_c\left(z+z_0\right)\sigma_z - p_z\sigma_x\right]$$
$$-\frac{\nu}{\hbar}\left[m^*\omega_c\left(z+z_0\right)\sigma_x - p_z\sigma_z\right]\}[\zeta(z)] = \varepsilon[\zeta(z)] \tag{7.41}$$

subject to the boundary condition $[\zeta(\infty)] = [\zeta(-\infty)] = [0]$. Here $z_0 = \hbar k_x/eB$ and $\omega_c = eB/m^*$.

Equation (7.41) has to be solved numerically to find the 2-component wavefunction $[\zeta(z)]$ and the energy dispersion relation ε versus k_x. Note that in the above equation $z_0 = \hbar k_x/eB$. However, since this equation is non-linear in z_0 (and hence k_x), a transformation of the type in Equation (7.11) will be required.

Note that if the y-directed external magnetic field is extremely strong, then obviously each element of the column vector $[\zeta(z)]$ will be the well known Landau orbit states $\mathcal{L}(z)$ which are simple harmonic oscillator wavefunctions centered at $z =$ -z_0 and for which the spatial averages $< z+z_0 > = < p_z > = 0$. In that case, the spin-orbit interaction terms in Equation (7.41) vanish. This is to be expected. The spin-orbit interaction actually gives rise to effective magnetic fields (see Chapter 8 for an exhaustive discussion). Therefore, if the external magnetic field is extremely strong, it will overwhelm the puny magnetic fields due to spin-orbit interaction and they will effectively vanish.

7.2 Dispersion relations of spin resolved magneto-electric subbands and eigenspinors in a one-dimensional electron gas in the presence of spin-orbit interaction

Consider a quantum wire, or one dimensional electron gas, as shown in Fig. 7.3. In contrast with the two-dimensional case, motions along both the y- and z-direction are constrained by quantum confinement and only motion along the x-direction is permitted. We will assume that the x-direction corresponds to the [100] crystallographic direction. There is a symmetry breaking electric field along the y-direction which induces the Rashba spin-orbit interaction. Furthermore, the quantum wire has crystallographic inversion asymmetry along the x-direction which induces the Dresselhaus spin-orbit interaction.

We will derive the dispersion relations and the eigenspinors in three different cases corresponding to an externally applied magnetic field being directed along three different coordinate axes - x, y and z.

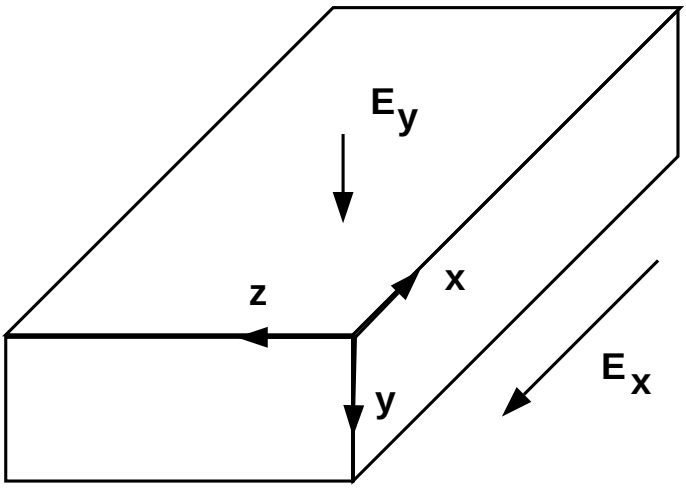


FIGURE 7.3
Schematic of a quantum wire showing the assignment of coordinate axes.

7.2.1 Magnetic field directed along the wire axis (x-axis)

The effective mass Hamiltonian for the wire, in the Landau gauge $\vec{A} = (0, -Bz, 0)$, can be written as

$$\begin{aligned} H^x_{1-DEG} = &(p_x^2 + p_y^2 + p_z^2)/(2m^*) - (eBzp_y)/m^* + (e^2B^2z^2)/(2m^*) \\ &-(g/2)\mu_B B\sigma_x + V(y) + V(z) \\ &+\nu_D[\sigma_x\kappa_x + \sigma_y\kappa_y + \sigma_z\kappa_z] - \eta[(p_x/\hbar)\sigma_z - (p_z/\hbar)\sigma_x], \end{aligned} \quad (7.42)$$

where $V(y)$ and $V(z)$ are the confining potentials along the y- and z-directions, ν is the strength of the Dresselhaus spin-orbit interaction and η is the strength of the Rashba spin-orbit Interaction, as before.

Since the Hamiltonian is invariant in the x-coordinate, the wavevector k_x is a good quantum number and the eigenstates are plane waves traveling in the x-direction. Therefore, the 2-component wavefunction can be written as

$$[\psi_{1-DEG}] = e^{ik_x x}\phi(y)[\lambda(z)], \quad (7.43)$$

where $[\lambda(z)]$ is a 2×1 column vector.

The Pauli Equation for this system is

$$H^x_{1-DEG}[\psi_{1-DEG}] = E[\psi_{1-DEG}]. \quad (7.44)$$

Spatially averaging both sides of the above equation over the x- and y-coordinates, we get

$$\begin{aligned} &[\{(\hbar^2k_x^2)/(2m^*) + E_m - \frac{\hbar^2}{2m^*}\frac{\partial^2}{\partial z^2} + \frac{e^2B^2z^2}{2m^*} + V(z)\}[I] \\ &\quad -(\nu k_x + \beta - i\eta\partial/\partial z)\sigma_x - (\eta k_x + i\nu\partial/\partial z)\sigma_z][\lambda(z)] = E[\lambda(z)], \end{aligned} \quad (7.45)$$

where $\beta = g\mu_B B/2$, ν = -$(\nu_D/\hbar^2) < p_y^2 > = \nu_D < \partial^2/\partial y^2 >$ and

$$\left[\frac{p_y^2}{2m^*} + V(y)\right]\phi_m(y) = E_m\phi_m(y). \quad (7.46)$$

In deriving the above, we assumed that the thickness of the quantum wire along the y-direction is much smaller than the width along the z-direction, so that $< p_y^2 > \; >> \; < p_z^2 >$ or $< p_x^2 >$. Note that since the y-component of the wavefunction is bounded, i.e., $\phi_m(y = d) = \phi_m(y = -d) = 0$, therefore $< p_y > = 0$.

In order to obtain the dispersion relation E versus k_x and the 2-component wavefunction $[\lambda(z)]$, we have to solve Equation (7.45) numerically, but only after transforming it following the trick of Equation (7.11) so that we first obtain an eigenequation in k_x. As always, this is the correct procedure, but it is also cumbersome and requires quite a bit of numerics. The approximate

technique detailed below allows us to obtain an analytical expression for the energy dispersion relation.

First, we will make the "variable separation" approximation, i.e., use the approximation in Equation (7.17) where we write the eigenspinor $[\lambda(z)]$ as the product of a space dependent part and a spin dependent part. We will assume that the spin-dependent part is independent of the z-coordinate. As a result, we will miss spin texturing effects where the spin density can vary along the z-coordinate. We have to pay this price in order to obtain analytical solutions. Second, we will assume that the energy separation between subbands (caused by confinement along the y- and z-directions) is so large that any mixing between them caused by spin-orbit interaction is negligible. This mixing can have interesting effects such as "anti-crossings" in the energy dispersion relations [2], but we will miss those effects as well. Finally, we will assume that the confinement along the z-direction is parabolic (this will indeed be the case if the confinement along this direction were imposed by "split-gates", as is usually done). In that case $V(z) = (1/2)m^*\omega_0^2 z^2$. Therefore, the unperturbed z-component of the wavefunction $\lambda_0(z)$ (in the absence of spin-orbit interaction) obeys the equation:

$$\left[\frac{p_z^2}{2m^*} + \frac{1}{2}m^*\omega^2 z^2\right]\lambda_0(z) = \Xi\lambda_0(z), \tag{7.47}$$

where $\omega^2 = \omega_0^2 + \omega_c^2$ and ω_c is the cyclotron frequency ($\omega_c = eB/m^*$).

In this case, the eigenenergies are $\Xi = (n + 1/2)\hbar\omega$, and $< p_z > = 0$. If we use the *unperturbed* wavefunction $\lambda_0(z)$ in Equation (7.45), and spatially average over the z-coordinate, then we obtain

$$E[\mathbf{I}] = \left[(\hbar^2 k_x^2)/(2m^*) + E_m + (n + 1/2)\hbar\omega\right][\mathbf{I}] - (\nu k_x + \beta)\sigma_x - (\eta k_x)\sigma_z. \tag{7.48}$$

The energy E is found from the two eigenvalues of the 2×2 matrix making up the right hand side of the above equation. This is the familiar "diagonalization" procedure and the two eigenvalues are

$$E_{\pm}^{(1)} = \frac{\hbar^2 k_x^2}{2m^*} + E_0 \pm \sqrt{(\eta^2 + \nu^2)\left(k_x + \frac{\nu\beta}{\eta^2 + \nu^2}\right)^2 + \frac{\eta^2}{\eta^2 + \nu^2}\beta^2}, \tag{7.49}$$

which are the energy-dispersion (E vs. k_x) relations for the two spin-split subbands. Here, $E_0 = E_m + (n + 1/2)\hbar\omega$ and we have used the superscript '(1)' in $E_{\pm}^{(1)}$ to remind the reader that this is the first of the three cases we will consider - where the magnetic field is along the x-axis (axis of the quantum wire).

The eigenspinors (2-component wavefunctions) are the eigenvectors of the matrix making up the right-hand side of the Equation (7.48) corresponding to the eigenvalues given in Equation (7.49). Therefore, these eigenspinors are

$$\Psi^{(1)}{}_{+}(B, k_x, x) = \begin{bmatrix} cos(\theta_{k_x}) \\ sin(\theta_{k_x}) \end{bmatrix} e^{ik_x x} \tag{7.50}$$

and

$$\Psi^{(1)}{}_{-}(B, k_x, x) = \begin{bmatrix} sin(\theta_{k_x}) \\ -cos(\theta_{k_x}) \end{bmatrix} e^{ik_x x}, \tag{7.51}$$

where $\theta_{k_x} = (1/2)arctan[(\nu k_x + \beta)/\eta k_x]$.

The dispersion relations given by Equation (7.49) are plotted in Fig. 7.4 (a) for the case $\eta = \nu$. Note that the dispersions are clearly nonparabolic and *asymmetric* about the energy-axis. More importantly, note that the eigenspinors given in Equation (7.51) are functions of k_x because θ_{k_x} depends on k_x. Therefore, the eigenspinors are not fixed in any spin-split level, in any subband, but change with k_x. In other words, neither level has a definite spin quantization axis and the spin of an electron in either level depends on the wavevector. We are familiar with this situation since we have seen it before in the case of the 2-DEG.

7.2.2 Spin components

Let us call the spin components along the x, y and z directions $S_x^{(1)\pm}$, $S_y^{(1)\pm}$ and $S_z^{(1)\pm}$, respectively. The $\pm$ sign refers to the two spin resolved levels. We will have:

$$S_m^{(1)\pm} = (\hbar/2)[\Psi_{\pm}^{(1)}]^{\dagger}[\sigma_m][\Psi_{\pm}^{(1)}] \quad m = x, y, z. \tag{7.52}$$

This yields $S_x^{(1)\pm}(k_x) = \pm\frac{\hbar}{2}sin(2\theta_{k_x})$, $S_y^{(1)\pm} = 0$ and $S_z^{(1)\pm}(k_x) = \pm\frac{\hbar}{2}cos(2\theta_{k_x})$.

It is easy to see that, in general, $S_x^{(1)}(k_x) \neq S_x^{(1)}(-k_x)$ and $S_z^{(1)}(k_x) \neq S_z^{(1)}(-k_x)$. However, if the Dresselhaus interaction vanishes ($\nu = 0$), then $S_x^{(1)}(k_x) = -S_x^{(1)}(-k_x)$ and $S_z^{(1)}(k_x) = S_z^{(1)}(-k_x)$. If the Rashba interaction vanishes, then $S_x^{(1)}(k_x) = \hbar/2$ and $S_y^{(1)}(k_x) = \mathrm{S}_z^{(1)}(k_x) = 0$.

7.2.2.1 Velocity versus wavevector relation

There is no velocity along y- and z-directions because of quantum confinement. The velocity along the x-direction is once again $\frac{\partial H}{\partial p_x}$:

$$\begin{aligned} v_x^{\pm} &= \frac{\hbar k_x}{m^*} \pm \frac{\eta}{\hbar} < \sigma_z > \pm \frac{\nu}{\hbar} < \sigma_x > \\ &= \frac{\hbar k_x}{m^*} \pm \frac{\eta}{\hbar} cos(2\theta_{k_x}) \pm \frac{\nu}{\hbar} sin(2\theta_{k_x}). \end{aligned} \tag{7.53}$$

As expected, the velocity is not proportional to k_x because of the spin-orbit interaction. Note that θ_{k_x} depends on k_x.

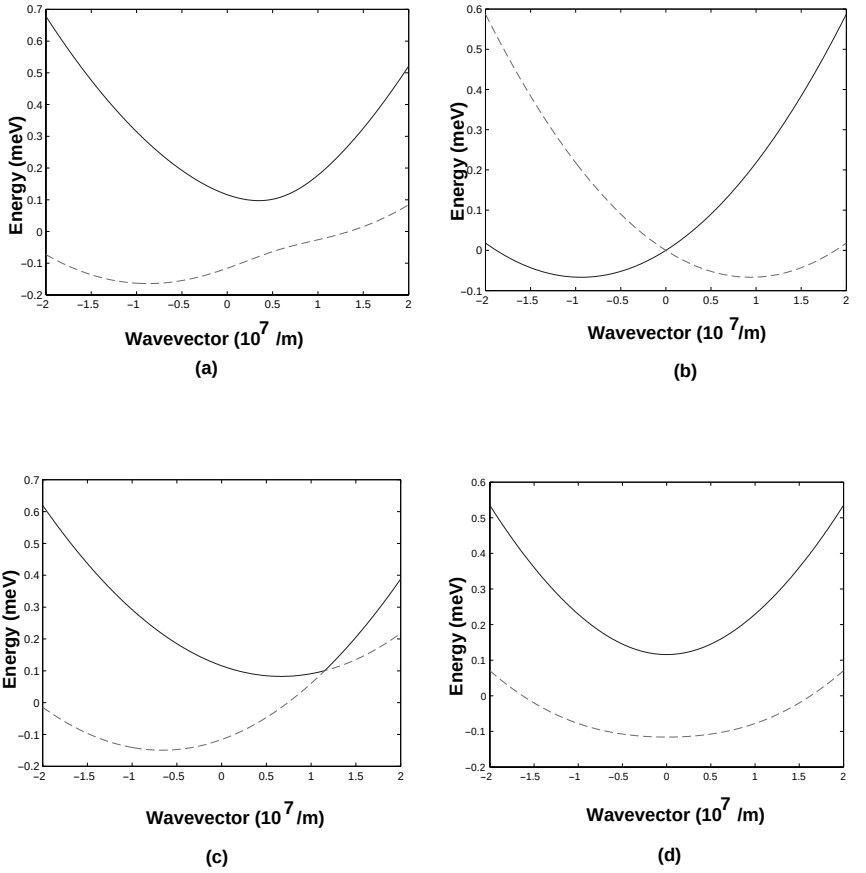


FIGURE 7.4
(a) Dispersion relations of spin-split subbands in a quantum wire with a magnetic field directed along the axis of the wire. We have plotted (assuming $E_0 = 0$) the dispersion relations for the cases: (a) $\eta = \nu = 10^{-11}$ eV-m, $m^* = 0.05m_0$, $B = 1$ Tesla, g = -2, (b) $\eta = \nu = 10^{-11}$ eV-m, $m^* = 0.05m_0$, $B = 0$ Tesla, (c) $\eta = 0$, $\nu = 10^{-11}$ eV-m, $m^* = 0.05m_0$, $B = 1$ Tesla, g = -2, and (d) $\eta = 10^{-11}$ eV-m, $\nu = 0$, $m^* = 0.05m_0$, $B = 1$ Tesla, and g = -2.

7.2.2.2 Special case

A special case arises when there is no magnetic field. In that case, the energy dispersion relation is

$$E_{\pm}^{(1)} = \frac{\hbar^2}{2m^*}\left(k_x \pm \frac{m^*\sqrt{\eta^2+\nu^2}}{\hbar^2}\right)^2 + E_0 - \frac{m^*(\eta^2+\nu^2)}{2\hbar^2}. \tag{7.54}$$

The above dispersion relation is definitely parabolic, albeit the origin of k_x has shifted by $\pm m^*\sqrt{\eta^2+\nu^2}/\hbar^2$ for the two spin-split bands. These dispersion relations are plotted in Fig. 7.4 (b). The role of spin-orbit interactions is to translate the dispersion parabolas (in any subband) horizontally from each other and shift them vertically downwards by a fixed amount. Note that the two parabolas are parallel to each other in the sense that at any energy, the difference between the wavevectors in the two subbands is $2m^*\sqrt{\eta^2+\nu^2}/\hbar^2$, *which is independent of the energy.* This feature plays a major role in the Spin Field Effect Transistor to be discussed in Chapter 12. In this case, θ_{k_x} is independent of k_x, so that the velocity is linearly related to the wavevector:

$$v_x^{\pm} = \frac{\hbar\,(k_x \pm constant)}{2m^*}. \tag{7.55}$$

Note that the velocity is zero when $k_x = \mp$ constant, but not when $k_x = 0$.

More importantly, if there is no magnetic field, then $\theta_{k_x} = (1/2)arctan[\nu/\eta]$, *independent* of k_x. In that case, the eigenspinors are wavevector-independent and each spin-split level in every subband has a definite spin quantization axis. The x-, y- and z-components of this spin quantization axis will be $S_x^{(1)\pm} = \pm\nu\hbar/(2\sqrt{\nu^2+\eta^2})$, $S_y^{(1)\pm} = 0$ and $S_z^{(1)\pm} = \pm\eta\hbar/(2\sqrt{\nu^2+\eta^2})$. The orientation of the spin quantization axis is determined by the relative strengths of the Dresselhaus and Rashba interactions alone since $S_x^{(1)\pm}/S_z^{(1)\pm} = \nu/\eta$. It does not depend on the wavevector.

Note that there is a crucial difference with the 2-DEG case. In the case of a 2-DEG, we needed that the strengths of the Rashba and Dresselhaus interactions be equal in order to make the eigenspinors independent of the wavevector. In the case of 1-DEG, we do *not* need the two strengths to be equal. All that we need to make the eignespinors independent of the wavevector is that there should be no magnetic field Absence of a magnetic field guarantees that each spin-split level has a fixed spin quantization axis.

An ultra-sensitive magnetometer

The above feature has an important device application. Whenever the eigenspinors are wavevector dependent, any momentum relaxing scattering event that changes the electron's wavevector – even if it is caused by a non-magnetic scatterer – will alter spin polarization and therefore cause spin relaxation. Now consider a device consisting of a quantum wire with two ideal ferromagnetic contacts that inject and detect spin with 100% efficiency. These

two ferromagnetic contacts are magnetized *anti-parallel* to each other. One contact injects electrons with a spin polarization that lies in the x-z plane and subtends an angle $arctan(\nu/\eta)$ with the z-axis. If there is no external magnetic field, then this spin is an eigenstate in the quantum wire. Therefore, unless there are magnetic scatterers to flip the spin, this spin will travel through the wire without flipping, even if there is plenty of momentum relaxing collisions that frequently change the electrons wavevector[†]. When the electron arrives at the other contact, it is completely blocked since the other contact is magnetized anti-parallel to the first contact, i.e., anti-parallel to the electrons spin. As a result, the current through the quantum wire will be ideally zero at any applied voltage. In other words, the wire resistance is infinite.

Now place this wire in a magnetic field along the wire axis. Immediately, the injected spins are no longer eigenstates in the wires since the eigenspinors become wavevector dependent. Any momentum relaxing scattering that changes the electron's wavevector will now also change the spin polarization. Consequently, the spins that arrive at the second contact will no longer necessarily be aligned anti-parallel to the magnetization of this contact. Therefore, there will be some current flow and the resistance of the wire will be finite. Consequently, this device can be used as a spintronic magnetic field sensor. It has infinite resistance in the absence of a magnetic field and a finite resistance in a magnetic field. A large array of such sensors (e.g., an array of $10^6 \times 10^6$ wires) can be extremely sensitive and with proper electronics detect magnetic fields with a sensitivity of a few tens of femto-Tesla/$\sqrt{Hz}$ at room temperature [4].

Obviously such a sensor can be realized *only with a quantum wire*. It cannot be realized with a quantum well (2-DEG) since the eigenspinors in a 2-DEG are not necessarily wavevector independent even if there is no magnetic field.

7.2.2.3 Another special case

Another special case arises if the Rashba interaction is absent. Then, $\theta_{k_x} = 45^\circ$, independent of k_x, even if there is a magnetic field. Each subband has a fixed spin quantization axis such that $S_y^{(1)\pm} = S_z^{(1)\pm} = 0$, and $S_x^{(1)\pm} = \pm\hbar/2$. The spin aligns itself parallel or anti-parallel to the direction of the *net* magnetic field which is the vector sum of the external magnetic field and the pseudo-magnetic field caused by the Dresselhaus interaction. Note that both these two magnetic fields are along the x-direction and therefore the spin lines up along the x-axis.

The reader can easily verify that the dispersion relation is again parabolic

[†]Basically, the two primary spin relaxation mechanisms - the Elliott-Yafet and Dyakonov-Perel (both discussed in the next chapter) – are absent.

in this case and given by

$$E_{\pm}^{(1)} = \frac{\hbar^2 \left(k_x \pm \frac{m^*\nu}{\hbar^2}\right)^2}{2m^*} + E_0 - \frac{m^*\nu^2}{2\hbar^2} \mp \beta. \tag{7.56}$$

The two parabolas representing the $E^{(1)}$ versus k_x relations of the two spin-split levels are now not only horizontally displaced from each other, but also vertically displaced. Again, the velocity is linearly related to the wavevector:

$$v_x^{\pm} = \frac{\hbar k_x}{m^*} \pm \frac{\nu}{\hbar}. \tag{7.57}$$

The dispersion relations for this case are shown in Fig. 7.4 (c).

7.2.2.4 Yet another special case

The last special case that we will consider is when the Dresselhaus interaction is absent. In this case, θ_{k_x} is wavevector dependent if there is a magnetic field; otherwise, $\theta_{k_x} = 0$, regardless of k_x. It is easy to show that in the latter case, $S_x^{(1)\pm} = S_y^{(1)\pm} = 0$ and $S_z^{(1)\pm} = \pm\hbar/2$. Therefore, the spin aligns itself parallel or anti-parallel to the direction of the pseudo magnetic field caused by the Rashba interaction.

7.2.3 Magnetic field perpendicular to wire axis and along the electric field causing Rashba effect (i.e., along y-axis)

Using the Landau gauge $\vec{A} = (Bz, 0, 0)$, the Hamiltonian for the wire can be written as

$$\begin{aligned} H_{1-DEG}^{y} = {} & (p_x^2 + p_y^2 + p_z^2)/(2m^*) + (eBzp_x)/m^* + (e^2B^2z^2)/(2m^*) \\ & -(g/2)\mu_B B\sigma_y + V(y) + V(z) \\ & +\nu_D[\sigma_x\kappa_x + \sigma_y\kappa_y + \sigma_z\kappa_z] \\ & -(\eta/\hbar)[(p_x + eBz)\sigma_z - p_z\sigma_x]. \end{aligned} \tag{7.58}$$

Since the Hamiltonian is invariant in the coordinate x, the wavefunctions are plane waves traveling in the x-direction, and can be written as

$$[\psi(x, y, z)] = e^{ik_x x}\phi_m(y)[\lambda(z)], \tag{7.59}$$

where $\phi_m(y)$ obeys the equation

$$\left[\frac{p_y^2}{2m^*} + V(y)\right]\phi_m(y) = E_m\phi_m(y). \tag{7.60}$$

By now, the reader should be able to determine that if we assume that the thickness of the quantum wire along the y-direction is much smaller than the

effective thickness along the z-direction, then the Pauli Equation will yield

$$\begin{aligned} E[\lambda(z)] = \{&[(\hbar^2 k_x^2 - \hbar^2(\partial^2/\partial z^2))/(2m^*) + (eBz\hbar k_x)/m^* + (e^2B^2z^2)/(2m^*) \\ &E_m + V(z)][I] - (g/2)\mu_B B\sigma_y \\ &-\nu\,[(k_x + eBz/\hbar)\sigma_x + i(\partial/\partial z)\sigma_z] \\ &-\eta\,[k_x + eBz/\hbar)\sigma_z + i(\partial/\partial z)\sigma_x]\}[\lambda(z)] \end{aligned} \tag{7.61}$$

subject to the boundary condition

$$[\lambda(z = \infty)] = [\lambda(z = -\infty)] = [0]. \tag{7.62}$$

We have to solve the above two equations, using the transformation trick of Equation (7.11), to find the energy dispersion relation E versus k_x, as well as the two-component wavefunction $[\lambda(z)]$. In Fig. 7.5, we show the dispersion relations calculated in this way for a quantum wire for two different values of the magnetic field. Then, in Fig. 7.6, we show the real and imaginary parts of the 2-component wavefunction $[\lambda(z)]$ $(= [\xi_\alpha(z), \xi_\beta(z)])$ as a function of the z-coordinate at a high magnetic field in two spin-split bands for a specific electron energy E_0. Note that the wavefunctions are skewed to one or the other edge of the wire depending on whether their velocities are positive or negative. This happens because the magnetic field exerts a Lorentz force $e\vec{v} \times \vec{B}$ on the electron, which pushes them towards one or the other edge of the wire depending on the direction of the velocity. Consequently, the wavefunctions are skewed towards one or the other edge of the quantum wire. These wavefunctions are therefore called "edge states" and play a critical role in the celebrated integer Quantum Hall Effect (as well as many other phenomena).

Since the Rashba and Dresselhaus spin-orbit interaction strengths vanish at $k_x = 0$, we would normally expect that the energy splitting between spin split bands at $k_x = 0$ should be the bare Zeeman splitting $g\mu_B B$. However, it turns out that this is *not* the case. Note that in Fig. 7.5, the spin-splitting energy at $k_x = 0$ clearly increases with increasing subband index, which should not have happened if it were pure Zeeman splitting since $g\mu_B B$ does not depend on the subband index. Although not obvious from the figure, it also turns out that the spin splitting energy within any subband is not even linear in the magnetic field, showing once again that it is not simply $g\mu_B B$. It would have been *approximately* $g\mu_B B$ had we made the variable separation approximation. The reason why this happens is because the spin-orbit interaction has an indirect influence on the so-called zero-k spin splitting. This has been discussed in ref. [5].

In order to obtain analytical results, we can make the variable separation approximation (recall Equation (7.17)) and neglect mixing between subbands caused by spin-orbit interaction. As mentioned earlier, we will then miss spin texturing effects and anti-crossing effects in the dispersion relation, but that is the price we have to pay in order to produce analytical results. Following

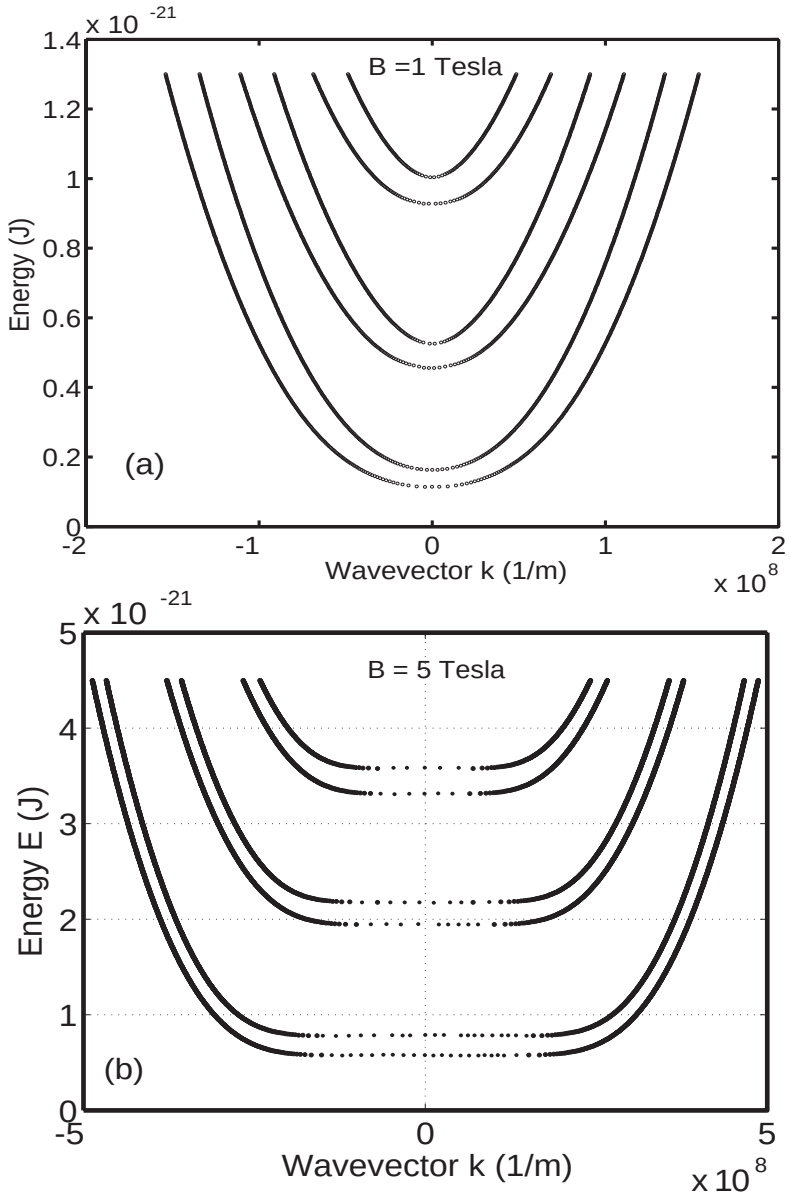


FIGURE 7.5

Energy dispersion relation ($E - k$ plot) in a quantum wire of width 100 nm in the presence of Rashba and Dresselhaus spin-orbit interaction and an external magnetic field in the y-direction. The effective mass is assumed to be 0.067 times the free electron mass and the Landé g-factor is assumed to be 4. The Rashba and Dresselhaus interaction strengths are $\eta = 2\nu = 10^{-11}$ eV-m. The magnetic flux density B is: (a) 1 Tesla, and (b) 5 Tesla. Reprinted with permission from Sandipan Pramanik, et al., Phys. Rev. B, **76**, 155325 (2007). Copyright (2007) American Physical Society. http://link.aps.org/abstract/PRB/v76/p155325.

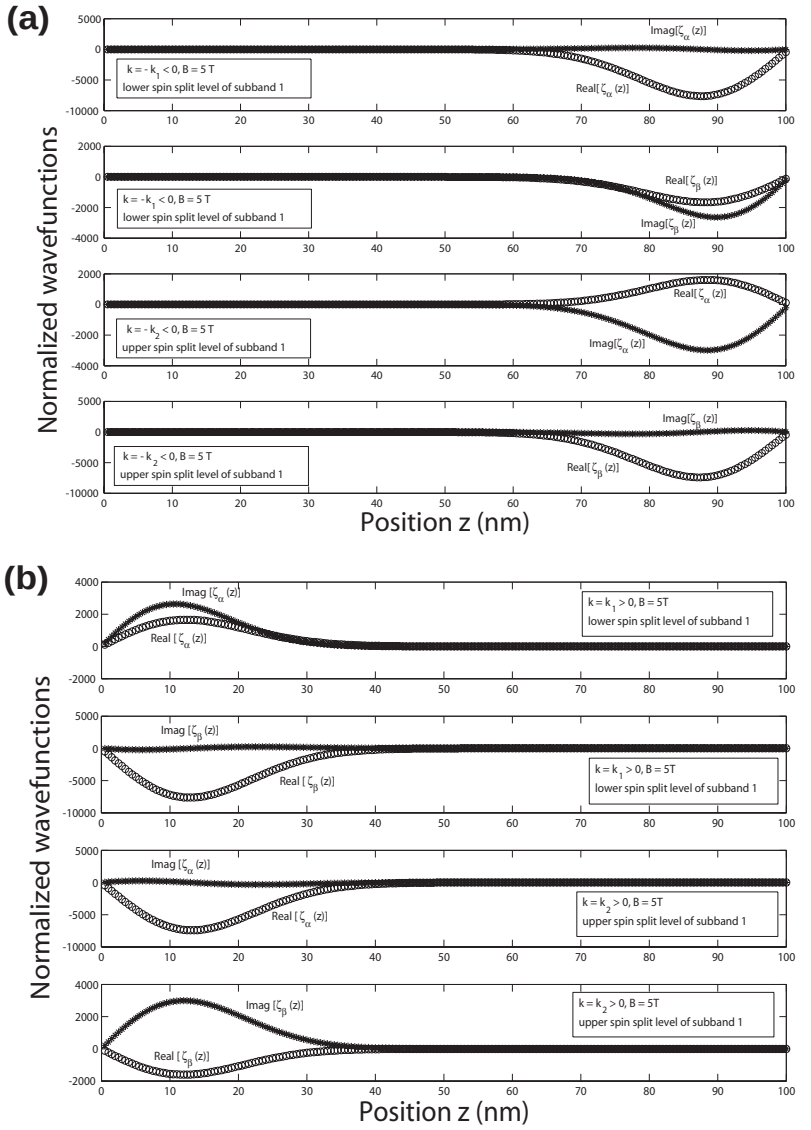


FIGURE 7.6

Real and imaginary parts of the two components of the wavefunction $[\zeta_\alpha(z) \;\; \zeta_\beta(z)]^T$ in a quantum wire as a function of the z-coordinate. The wavefunctions are plotted for electron energy $E_0 = 9.3$ meV and magnetic flux density $B = 5$ Tesla. The wavevectors corresponding to this energy in the two spin-split bands are $k_1^{\pm} = \pm\, 3.56 \times 10^8$ m^{-1} (lower band), and $k_2^{\pm} = \pm\, 3.3 \times 10^8$ m^{-1} (upper band). In Fig. (a), we show the wavefunctions for k_1^- and k_2^-, while in Fig. (b), we show the wavefunctions for k_1^+ and k_2^+. The wavefunctions are skewed to the left or right edge of the quantum wire, depending on whether the electrons are forward or backward traveling (i.e., whether k is positive or negative), because of the Lorentz force that pushes electrons towards one or the other edge.

the usual procedure (i.e., spatially averaging the Hamiltonian over x-, y- and z-coordinates) and then carrying out the "diagonalization", we obtain the energy dispersion relations of the two spin-split levels as

$$E_{\pm}^{(2)} = \frac{\hbar^2 k_x^2}{2m^*} + \epsilon_n + \left(m + \frac{1}{2}\right)\hbar\omega \pm \sqrt{P^2 - Q^2 + \left(g\mu_B B/2\right)^2}, \tag{7.63}$$

where

$$\begin{aligned} P^2 &= \left(\nu^2 + \eta^2\right)\left[\left(k_x + eB < z > /\hbar\right)^2 + < p_z/\hbar >^2\right] \\ Q^2 &= 4(\nu\eta/\hbar^2)eB < z >< p_z > . \end{aligned} \tag{7.64}$$

Here we have used the superscript '(2)' in $E_{\pm}^{(2)}$ to remind the reader that we are now dealing with the second of the three cases, namely that the magnetic field is parallel to the y-axis.

We assumed that the confinement along the z-direction is parabolic so that the confining potential $V(z) = (1/2)m^*\omega_0^2$, $\omega^2 = \omega_0^2 + \omega_c^2$, $\omega_c = eB/m^*$, $< z > = \int_0^\infty dz\lambda_0^*(z)z\lambda_0(z)$, $< p_z > = \int_0^\infty dz\lambda_0^*(z)(-i\hbar(\partial/\partial z)\lambda_0(z)$ and ϵ_n is the quantized energy in the n-th subband due to confinement along the y-direction. Because of confinement along the z-direction, $< p_z > = 0$. Therefore, the above equation simplifies to

$$E_{\pm}^{(2)} = \frac{\hbar^2 k_x^2}{2m^*} + \epsilon_n + \left(m + \frac{1}{2}\right)\hbar\omega \pm \sqrt{\left(\nu^2 + \eta^2\right)\left(k_x + eB < z > /\hbar\right)^2 + \left(g\mu_B B/2\right)^2}. \tag{7.65}$$

In the absence of spin-orbit interaction, we can show that $(k_x + eB < z > /\hbar) = 0$. However, when spin-orbit interaction is present, $(k_x + eB < z > /\hbar) \neq 0$. Note that the spin-splitting at $k_x = 0$ is not the Zeeman splitting $g\mu_B B$, but rather $g^*\mu_B B$, where

$$g^* = g\sqrt{1 + \frac{4\left(\nu^2 + \eta^2\right)e^2 < z_0 >^2}{\hbar^2 g\mu_B}}, \tag{7.66}$$

where $< z_0 >$is the value of $< z >$ when $k_x = 0$[‡].

The eigenspinors are:

$$\Psi_{-}^{(2)}(B, k_x) = \begin{bmatrix} cos(\theta'_{k_x}) \\ sin(\theta'_{k_x})e^{-i\phi_{k_x}} \end{bmatrix} \tag{7.67}$$

$$\Psi_{+}^{(2)}(B, k_x) = \begin{bmatrix} sin(\theta'_{k_x}) \\ -cos(\theta'_{k_x})e^{-i\phi_{k_x}} \end{bmatrix}, \tag{7.68}$$

[‡]Actually $< z_0 >$ may also have a slight dependence on the magnetic flux density B, but we have ignored that here.

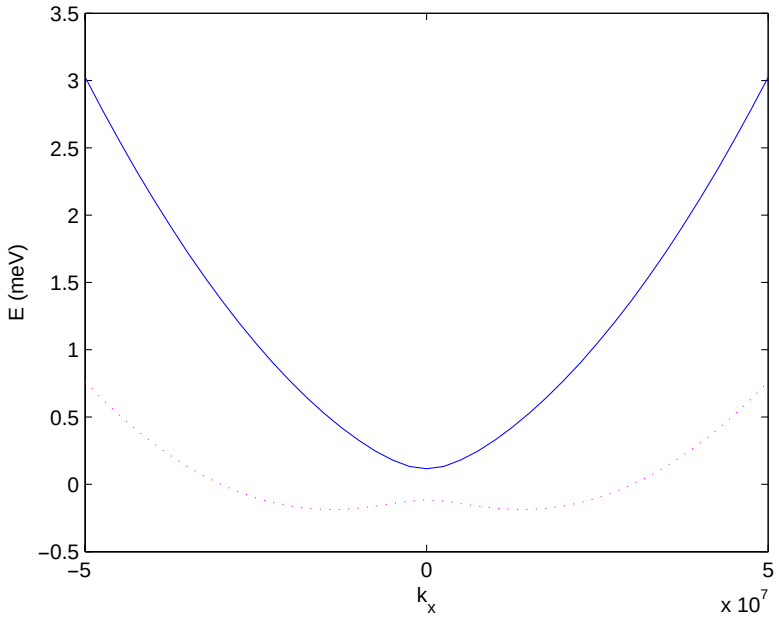


FIGURE 7.7
Dispersion relations of spin-split subbands in a quantum wire with a magnetic field directed perpendicular to the axis of the wire along the direction in which there is a symmetry breaking electric field causing a Rashba interaction. We have assumed that $< z > = 0$, $m^* = 0.05m_0$, $\eta = \nu = 10^{-11}$ eV-m, $g = 4$ and $B = 1$ Tesla.

where

$$\theta'_{k_x} = -(1/2)arctan[\{ig\mu_B B/2+\nu(k_x+eB < z > /\hbar)\}/\eta(k_x + eB < z > /\hbar)]. \tag{7.69}$$

Note that the angle θ'_{k_x} is a complex quantity. We leave it to the reader to infer what is the angle ϕ_{k_x}.

In Fig. 7.7, we plot the dispersion relations (E versus k_x) for the case $\eta = \nu$. Note that the dispersions are again clearly nonparabolic. The eigenspinors are functions of k_x because θ'_{k_x} and ϕ_{k_x} depend on k_x. Therefore, the eigenspinors are again not fixed in any subband, but change with k_x. Neither subband has a definite spin quantization axis and the spin of an electron in either subband depends on the wavevector.

7.2.4 Spin components

The spin components along the x, y and z directions are given by

$$S_m^{\pm} = (\hbar/2)[\Psi_{\pm}^*]^{\dagger}[\sigma_m][\Psi_{\pm}] \quad m = x, y, z. \tag{7.70}$$

This yields

$$\begin{aligned} S_x^{\pm}(k_x) &= \pm\hbar Re[cos^*(\theta'_{k_x})sin(\theta'_{k_x})], \\ S_y^{\pm}(k_x) &= \pm\hbar Im[cos^*(\theta'_{k_x})sin(\theta'_{k_x})], \\ S_z^{\pm}(k_x) &= \pm(\hbar/2)\left[|cos(\theta'_{k_x})|^2 - |sin(\theta'_{k_x})|^2\right]. \end{aligned} \tag{7.71}$$

Here 'Re' stands for the real part and 'Im' stands for the imaginary part.

7.3 Magnetic field perpendicular to wire axis and the electric field causing Rashba effect (i.e., along the z-axis)

Using the Landau gauge $\vec{A} = (-By,\, 0,\, 0)$, the Hamiltonian for the wire can be written as

$$\begin{aligned} H_{1-DEG}^z = {} & (p_x^2 + p_y^2 + p_z^2)/(2m^*) - (eByp_x)/m^* + (e^2B^2y^2)/(2m^*) \\ & -(g/2)\mu_B B\sigma_z + V(y) + V(z) \\ & +\nu_D[\sigma_x\kappa_x + \sigma_y\kappa_y + \sigma_z\kappa_z] \\ & -(\eta/\hbar)[(p_x - eBy)\sigma_z - p_z\sigma_x]. \end{aligned} \tag{7.72}$$

The 2-component wavefunction which is a solution of the above Hamiltonian can be written as

$$[\psi(x, y, z)] = e^{ik_x x}\lambda(z)[\phi(y)], \tag{7.73}$$

where the z-component of the unperturbed wavefunction (in the absence of spin-orbit interaction) obeys the equation

$$\left[\frac{p_z^2}{2m^*} + V(z)\right]\lambda_0(z) = \varepsilon\lambda_0(z). \tag{7.74}$$

If the confinement along the z-direction is parabolic, i.e., $V(z) = (1/2)m^*\omega_0^2 z^2$, then the wavefunctions $\lambda_0^n(z)$ for the n-th subband are simple harmonic oscillator wavefunctions:

$$\lambda_0^n(z) = \left(\frac{\Gamma'}{\sqrt{\pi}2^n n!}\right)^{1/2} \mathcal{H}_n(\Gamma' z)e^{-\frac{1}{2}(\Gamma')^2 z^2}, \tag{7.75}$$

where $\Gamma' = \{(m^*\omega_0)^2/(2\hbar^2)\}^{1/4}$ and $\mathcal{H}_n$ is the Hermite polynomial of the n-th order.

As before, we will consider two cases: (i) the magnetic field is so strong that the magnetostatic confinement along the y-direction is much stronger than then electrostatic confinement $V(y)$, and (ii) the opposite case, namely that the electrostatic confinement is stronger than the magnetostatic confinement.

7.3.0.1 Case I: Magnetostatic confinement is stronger

In this case, the y-component of the unperturbed wavefunction (in the absence of spin-orbit interaction) is given by the simple harmonic oscillator states

$$\phi_0^m(y) = \left(\frac{\Gamma}{\sqrt{\pi}2^m m!}\right)^{1/2} \mathcal{H}_m(\Gamma[y-y_0]) e^{-\frac{1}{2}\Gamma^2(y-y_0)^2}, \tag{7.76}$$

where $\Gamma = \{(m^*\omega_c)^2/(2\hbar^2)\}^{1/4}$ and $\mathcal{H}_m$ is the Hermite polynomial of the m-th order, and $y_0 = \hbar k_x/eB$.

It is easy to show that the expectation value of the Hamiltonian in Equation (7.72) (evaluated using the unperturbed wavefunction) is

$$\begin{aligned} &\langle \lambda_0^n(z)\phi_0^m(y)e^{ik_x x}|H^z_{1-DEG}|\lambda_0^n(z)\phi_0^m(y)e^{ik_x x}\rangle \\ &= E\mathbf{I} = [(n+1/2)\hbar\omega_0 + (m+1/2)\hbar\omega_c]\mathbf{I} - \beta\sigma_z. \end{aligned} \tag{7.77}$$

In deriving the above, we have used the fact that $< p_x - eBy > = 0$.

As usual, the energy eigenstates are found by diagonalization, i.e., they are the eigenvalues of the 2×2 matrix making up the right hand side of the above equation. They are:

$$E_{\pm}^{(3)} = (n+1/2)\hbar\omega_0 + (m+1/2)\hbar\omega_c \pm \beta \tag{7.78}$$

where the superscript '(3)' in $E_{\pm}^{(3)}$ is intended to remind the reader that we are dealing with the last of the three cases, when the magnetic field is in the z-direction. The last equation gives the energies of the two spin split levels (independent of k_x) and the corresponding eigenspinors are

$$\Psi_+^{(3)}(B, k_x, x) = \begin{bmatrix} 1 \\ 0 \end{bmatrix} e^{ik_x x} \tag{7.79}$$

$$\Psi_-^{(3)}(B, k_x, x) = \begin{bmatrix} 0 \\ 1 \end{bmatrix} e^{ik_x x}. \tag{7.80}$$

Note that there is no influence of spin-orbit interaction in this case. Only the Zeeman interaction splits the spin eigenstates. This is a familiar situation. The spin-orbit interactions will cause effective magnetic fields which are completely overwhelmed by the externally applied magnetic field because it is much stronger. In that case, we see no effect of the spin-orbit interaction.

7.3.0.2 Case II: Electrostatic confinement is stronger

In this case, let the y-component of the wavefunction be $\phi_m(y)$ and let $< y > = < \phi_m(y)|y|\phi_m(y) > = Y_m$ and $< y^2 > = < \phi_m(y)|y^2|\phi_m(y) > = (\mathcal{Y}_m)^2$. Therefore, the expectation value of the Hamiltonian in Equation (7.72) is

$$\begin{aligned}\langle H^z_{1-DEG}\rangle = E[\mathbf{I}] = &\left[\frac{\hbar^2 k_x^2}{2m^*} + \epsilon_m + (n+1/2)\hbar\omega_0 - \frac{eBY_m\hbar k_x}{m^*} + \frac{e^2B^2(\mathcal{Y}_m)^2}{2m^*}\right][\mathbf{I}] \\ &-\beta\sigma_z - \nu[k_x - eBY_m/\hbar]\sigma_x \\ &-\eta[k_x - eBY_m/\hbar]\sigma_z \\ = &\left[\frac{\hbar^2 k_x^2}{2m^*} + \varepsilon_{m,n} - \frac{\hbar^2\Delta}{m^*}k_x\right]\mathbf{I} - \beta\sigma_z \\ &-\nu(k_x-\Delta)\sigma_x - \eta(k_x-\Delta)\sigma_z, \end{aligned} \tag{7.81}$$

where ϵ_m are the energies of the subbands caused by the confinement in the y-direction, $\varepsilon_{m,n} = \epsilon_m + (n+1/2)\hbar\omega_0 + \frac{e^2B^2(\mathcal{Y}_m)^2}{2m^*}$ and $\Delta = eBY_m/\hbar$. In deriving the above, we used the fact that the spatial averages $< p_y > = < p_z > = 0$.

The energy eigenstates are the eigenvalues of the 2×2 matrix making up the right hand side of the above equation. These eigenenergies are:

$$E^{(3)}_{\pm} = \frac{\hbar^2 k_x^2}{2m^*} + \varepsilon_{m,n} - \frac{\hbar^2\Delta k_x}{2m^*} \pm \sqrt{(\eta^2+\nu^2)\left(k_x-\Delta\right)^2 + 2\eta\beta\left(k_x-\Delta\right) + \beta^2}, \tag{7.82}$$

and the corresponding eigenspinors are

$$\Psi^{(3)}{}_+(B,k_x,x) = \begin{bmatrix} cos(\theta''_{k_x}) \\ sin(\theta''_{k_x}) \end{bmatrix} e^{ik_x x} \tag{7.83}$$

$$\Psi^{(3)}{}_-(B,k_x,x) = \begin{bmatrix} sin(\theta''_{k_x}) \\ -cos(\theta''_{k_x}) \end{bmatrix} e^{ik_x x}, \tag{7.84}$$

where $\theta''_{k_x} = (1/2)arctan[\{\nu(k_x-\Delta)\}/\{\eta(k_x-\Delta)+\beta\}]$.

The eigenspinors are functions of k_x because θ''_{k_x} depends on k_x. Therefore, the eigenspinors are again not fixed in any subband, but change with k_x. Neither subband has a definite spin quantization axis and the spin of an electron in either subband depends on the wavevector.

The reader should note that in every dispersion relation that we have derived, the quantity under the radical ("square root sign") always remains positive at any value of the wavevector k_x. That means there is no value of k_x at which the energy becomes a complex quantity which would have been unphysical.

7.3.1 Spin components

The spin components along the x, y and z directions are given by

$$S_m^{(3)\pm} = (\hbar/2)[\Psi_\pm^{(3)}]^\dagger[\sigma_m][\Psi_\pm^{(3)}] \quad m = x, y, z. \tag{7.85}$$

This yields $S_x^{(3)\pm}(k_x) = \pm(\hbar/2)sin(2\theta''_{k_x})$, $S_y^{(3)\pm} = 0$ and $S_z^{(3)\pm}(k_x) = (\hbar/2) \pm cos(2\theta''_{k_x})$.

In this case, $S_x^{(3)}(k_x) \neq S_x^{(3)}(-k_x)$ and $S_z^{(3)}(k_x) \neq S_z^{(3)}(-k_x)$ unless the Dresselhaus interaction vanishes so that $\nu = \Delta = 0$. In the latter case, $S_x^{(3)}(k_x) = -S_x^{(3)}(-k_x)$ and $S_z^{(3)}(k_x) = S_z^{(3)}(-k_x)$.

7.3.2 Special case

A special case arises if the Dresselhaus interaction is absent. In this case, it is easy to show that

$$E_\pm^{(3)} = \frac{\hbar^2}{2m^*}\left[k_x - \left(\frac{\Delta}{2} \pm \frac{m^*\eta}{\hbar^2}\right)\right]^2 + \varepsilon_{m,n} - \frac{\hbar^2}{2m^*}\left(\frac{\Delta}{2} \pm \frac{m^*\eta}{\hbar^2}\right)^2 \mp \eta\Delta \pm \beta \tag{7.86}$$

In this case, the dispersion relations are parabolic. The two spin resolved parabolas are horizontally displaced from each other by a fixed wavevector $2\frac{m^*\eta}{\hbar^2}$ and vertically displaced from each other by a fixed energy.

7.4 Eigenenergies of spin resolved subbands and eigenspinors in a quantum dot in the presence of spin-orbit interaction

Consider a quantum dot, or quasi zero-dimensional electron gas, as shown in Fig. 7.8. Here, motion along every direction is constrained by quantum confinement. There is a symmetry breaking electric field along the y-direction which, under proper conditions that we will enumerate later, induces the Rashba spin-orbit interaction. Furthermore, the quantum dot has crystallographic inversion asymmetry which induces the Dresselhaus spin-orbit interaction.

As always, we will consider scenarios when there is an external magnetic field. Let us assume that the magnetic field is directed along the x-direction. We choose the Landau gauge $\vec{A}$ = -$Bz\hat{y}$.

The effective mass Hamiltonian describing this system is

$$H_{dot} = \frac{p_x^2 + (p_y - eBz)^2 + p_z^2}{2m^*} + V_x(x) + V_y(y) + V_z(z)$$

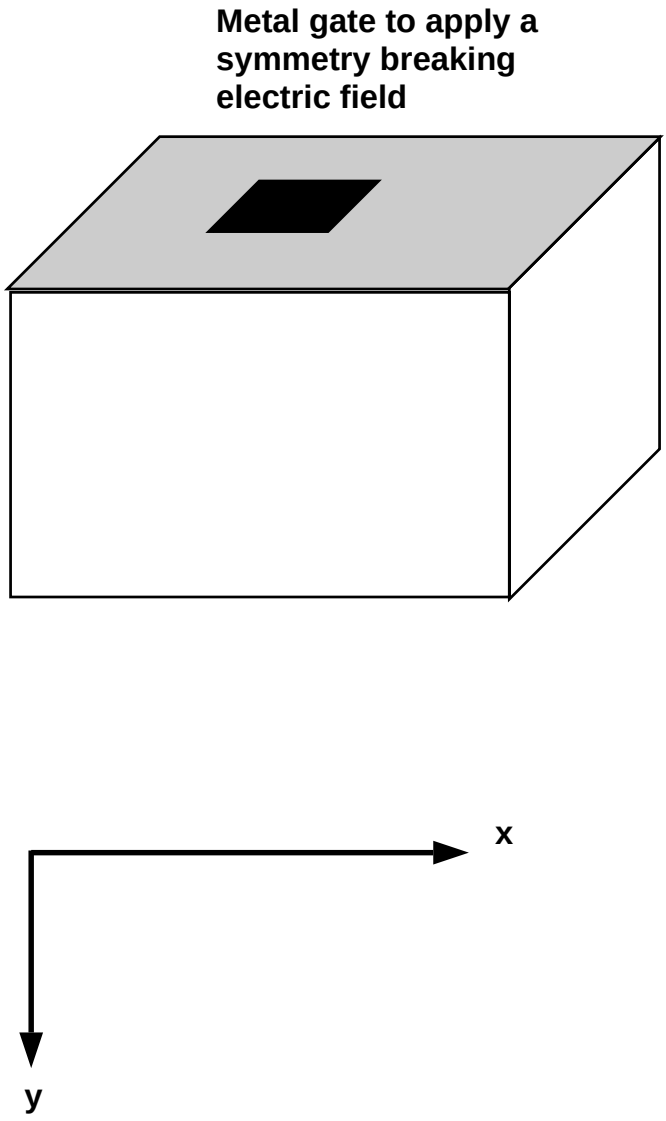


FIGURE 7.8
Schematic representation of a rectangular quantum dot with spin orbit interaction placed in a magnetic field.

$$
\begin{aligned}
&-(g/2)\mu_B B\sigma_x - \frac{\eta}{\hbar}\left[p_x\sigma_z - p_z\sigma_x\right] \\
&+\frac{\nu}{2\hbar^3}\left[p_x\{(p_y - eBz)^2 - p_z^2\} + \{(p_y - eBz)^2 - p_z^2\}p_x\right]\sigma_x \\
&+\frac{\nu}{2\hbar^3}\left[(p_y - eBz)\{p_z^2 - p_x^2\} + \{p_z^2 - p_x^2\}(p_y - eBz)\right]\sigma_y \\
&+\frac{\nu}{2\hbar^3}\left[p_z\{p_x^2 - (p_y - eBz)^2\} + \{p_x^2 - (p_y - eBz)^2\}p_z\right]\sigma_z, \quad (7.87)
\end{aligned}
$$

where $V_x(x)$, $V_y(y)$ and $V_z(z)$ are the confining potentials along the three coordinate axes, and $V_y(y)$ includes the potential due to the electric field inducing the Rashba interaction.

In a quantum dot, spin-orbit interaction is weak. Normally, the strengths of the spin-orbit interactions are proportional to the momentum. A quantum dot has no extended states and the expected value of the momentum operator $< p_x >, < p_y >$ or $< p_z >$ is zero. We will show that because of this, there can be no effect of spin-orbit interaction on eigenenergies and eigenstates, *unless there is a magnetic field* which lifts the spin degeneracy via the Zeeman effect, and in the process *makes the spatial parts of the wavefunctions of the spin resolved states different if the quantum dot has finite potential barriers.* Spin-orbit interaction can then couple different subbands and have a non-trivial effect.

In order to find the eigenstates of Equation (7.87), we will apply perturbation theory. The unperturbed wavefunctions are those without the spin orbit interactions. These 2-component wavefunctions (in the lowest Zeeman split subband) are

$$
\begin{aligned}
\{\phi_\uparrow(x,y,z)\} &= \frac{1}{\sqrt{2}}\phi_x(x)\phi_y(y)\phi_z(z)\begin{bmatrix}1\\1\end{bmatrix} \\
\{\phi_\downarrow(x,y,z)\} &= \frac{1}{\sqrt{2}}\phi'_x(x)\phi'_y(y)\phi'_z(z)\begin{bmatrix}1\\-1\end{bmatrix},
\end{aligned} \quad (7.88)
$$

where we have assumed that the spatial parts of the wavefunctions of the two spin resolved states (which are resolved by the Zeeman effect) are *different.* This will happen as long as the quantum dot has finite potential barriers. In that case, the upper spin level's wavefunction will be less confined than the lower spins level's wavefunction and therefore will have a larger spatial spread, making $\phi \neq \phi'$. Note that the magnetic field is essential for this inequality and without this inequality, there will be no Rashba or Dresselhaus spin-orbit interaction in the dot, at least to first order.

Note that the unperturbed eigenspinors (unperturbed by spin-orbit interaction) are +x and -x-polarized spin states $(1/\sqrt{2})[1\ 1]^\dagger$ and $(1/\sqrt{2})[1\ -1]^\dagger$ since the external magnetic field is in the x-direction.

The Pauli Equation is

$$
[H_{dot}]\,[\psi] = E[\psi], \quad (7.89)
$$

where the 2-component wavefunction $[\psi]$ is a linear superposition of "up-spin" and "down-spin" states, i.e., +x- and -x-polarized states, that are mutually orthogonal:

$$[\psi] = a_{\uparrow}\{\phi^{\uparrow}\} + a_{\downarrow}\{\phi^{\downarrow}\}. \tag{7.90}$$

Equation (7.89) can be re-written as

$$[H_0 + H_{SO}]\,[\psi] = E[\psi], \tag{7.91}$$

where H_0 is the unperturbed Hamiltonian and H_{SO} is the spin-orbit interaction Hamiltonian due to the Rashba and Dresselhaus interactions.

Substitution of Equation (7.90) into Equation (7.91) yields

$$\begin{bmatrix} < H_1 > + < H_{SO} >_{11} & < H_{SO} >_{12} \\ < H_{SO} >_{21} & < H_2 > + < H_{SO} >_{22} \end{bmatrix} \begin{pmatrix} a_{\uparrow} \\ a_{\downarrow} \end{pmatrix} = E_n \begin{pmatrix} a_{\uparrow} \\ a_{\downarrow} \end{pmatrix}, \tag{7.92}$$

where $< H_1 > = < \phi_s^{\uparrow}|H_0|\phi_s^{\uparrow} >$, $< H_2 > = < \phi_s^{\downarrow}|H_0|\phi_s^{\downarrow} >$, $< H_{SO} >_{11} = < \phi_s^{\uparrow}|H_{SO}|\phi_s^{\uparrow} >$, $< H_{SO} >_{22} = < \phi_s^{\downarrow}|H_{SO}|\phi_s^{\downarrow} >$, $< H_{SO} >_{12} = < \phi_s^{\uparrow}|H_{SO}|\phi_s^{\downarrow} >$ and $< H_{SO} >_{21} = < \phi_s^{\downarrow}|H_{SO}|\phi_s^{\uparrow} >$. Here, $\phi_s^{\uparrow} = \phi_x(x)\phi_y(y)\phi_z(z)$ and $\phi_s^{\downarrow} = \phi'_x(x)\phi'_y(y)\phi'_z(z)$.

It is easy to show that $< H_{SO} >_{11} = < H_{SO} >_{22} = 0$. Thereafter, diagonalizing the above Hamiltonian (in the basis of the states given in Equation (7.88)), we get

$$E = \frac{< H_1 > + < H_2 >}{2} \pm \sqrt{\left(\frac{g\mu_B B}{2}\right)^2 + < H_{SO} >_{12} < H_{SO} >_{21}}, \tag{7.93}$$

where we have used the fact that $< H_1 > - < H_2 > = g\mu_B B$.

Let us first find the contribution to $< H_{SO} >_{12}$ due to Rashba interaction. The Rashba Hamiltonian is

$$H_{SO}^{Rashba} = -(\eta/\hbar)[p_x\sigma_z - p_z\sigma_x]. \tag{7.94}$$

Therefore,

$$\begin{aligned} < H_{SO}^{Rashba} >_{12} &= \frac{-\eta}{2\hbar} \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} -\overline{p_x} & \overline{p_z} \\ \overline{p_z} & \overline{p_x} \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} \\ &= \frac{\eta}{\hbar}\overline{p_x} \\ &= D, \end{aligned} \tag{7.95}$$

where $\overline{p_x} = < \phi_x(x)\phi_y(y)\phi_z(z)| - i\hbar(\partial/\partial x)|\phi'_x(x)\phi'_y(y)\phi'_z(z) >$. Note that $\overline{p_x} \neq 0$, only because the spatial wavefunctions of the Zeeman split states are *different* ($\phi_x(x) \neq \phi'_x(x)$). It is easy to show (following the same prescription as above) that

$$< H_{SO}^{Rashba} >_{21} = \frac{\eta}{\hbar}\overline{p'_x} = D^*, \tag{7.96}$$

where $\overline{p'_x} = < \phi'_x(x)\phi'_y(y)\phi'_z(z)| - i\hbar(\partial/\partial x)|\phi_x(x)\phi_y(y)\phi_z(z) >$. Since the operator $-i\hbar(\partial/\partial x)$ is Hermitean, one can shown that $\overline{p'_x}$ is the complex conjugate of $\overline{p_x}$, so that D^* is the complex conjugate of D. Furthermore, D is completely imaginary since the wavefunction ϕ_x is *real*. Therefore, D^* = -D.

The Dresselhaus Hamiltonian is

$$\begin{aligned} H_{SO}^{Dresselhaus} = & \frac{\nu}{2\hbar^3}\left[p_x\{(p_y - eBz)^2 - p_z^2\} + \{(p_y - eBz)^2 - p_z^2\}p_x\right]\sigma_x \\ & + \frac{\nu}{2\hbar^3}\left[(p_y - eBz)\{p_z^2 - p_x^2\} + \{p_z^2 - p_x^2\}(p_y - eBz)\right]\sigma_y \\ & + \frac{\nu}{2\hbar^3}\left[p_z\{p_x^2 - (p_y - eBz)^2\} + \{p_x^2 - (p_y - eBz)^2\}p_z\right]\sigma_z. \end{aligned} \tag{7.97}$$

Therefore,

$$\begin{aligned} < H_{SO}^{Dresselhaus} >_{12} &= \frac{\eta}{2\hbar}[\begin{bmatrix}1 & 1\end{bmatrix}\begin{bmatrix}A & B \\ C & -A\end{bmatrix}\begin{bmatrix}1 \\ -1\end{bmatrix} \\ &= A - (B - C)/2, \end{aligned} \tag{7.98}$$

where A, B and C are complex constants. The reader can evaluate these constants by evaluating the expectation values of the operators in Equation (7.97). It is easy to show that

$$< H_{SO}^{Dresselhaus} >_{21} = A' - (B' - C')/2, \tag{7.99}$$

where, once again, the reader can evaluate the complex constants A', B' and C' from the expectation values of the operators in the Hamiltonian of Equation (7.97).

It should be obvious that $< H_{SO}^{Dresselhaus} >_{12} \neq 0$ and $< H_{SO}^{Dresselhaus} >_{21} \neq 0$, only because the spatial wavefunctions of the Zeeman split states are *different*. Thus, without the magnetic field, there will be no spin-orbit interaction effect on a quantum dot's energy dispersion relation or eigenspinors.

The reader can also show that $< H_{SO}^{Dresselhaus} >_{12} = < H_{SO}^{Dresselhaus} >_{21}^*$ where the asterisk denotes complex conjugate. This is a consequence of the fact that the spin-orbit Hamiltonian must be Hermitean like all quantum mechanical operators. Therefore, $< H_{SO} >_{12} < H_{SO} >_{21} = | < H_{SO} >_{12} |^2$.

Finally, substituting all this in Equation (7.93), we get the eigenenergies in a quantum dot in the presence of spin-orbit interaction:

$$E_{dot} = \frac{< H_1 > + < H_2 >}{2} \pm \sqrt{\left(\frac{g\mu_B B}{2}\right)^2 + |\delta|^2}, \tag{7.100}$$

where

$$\delta = < H_{SO} >_{12} = [D + A - (B - C)/2]. \tag{7.101}$$

We can now proceed to find the coefficients $a_\uparrow$ and $a_\downarrow$, which are the eigenfunction of the matrix in Equation (7.92). For the lower spin-split state, corresponding to the negative sign in Equation (7.100),

$$\begin{aligned} a_\uparrow^- &= cos\zeta \\ a_\downarrow^- &= sin\zeta e^{i\xi}, \end{aligned} \tag{7.102}$$

where $\zeta = (1/2)\arctan[|\delta|/(g\mu_B B)]$ and ξ is the phase of δ, i.e., $\delta = |\delta|e^{i\xi}$. For the higher spin-split state,

$$\begin{aligned} a_\uparrow^+ &= -sin\zeta \\ a_\downarrow^+ &= cos\zeta e^{i\xi}. \end{aligned} \tag{7.103}$$

Therefore, the eigenspinor for the lower spin-split state is

$$\begin{aligned} \Psi_- &= \frac{a_\uparrow^-}{\sqrt{2}}\begin{bmatrix}1\\1\end{bmatrix} + \frac{a_\downarrow^-}{\sqrt{2}}\begin{bmatrix}1\\-1\end{bmatrix} \\ &= \frac{1}{\sqrt{2}}\begin{bmatrix}cos\zeta + sin\zeta e^{i\xi}\\ cos\zeta - sin\zeta e^{i\xi}\end{bmatrix}, \end{aligned} \tag{7.104}$$

and the eigenspinor in the higher spin-split state is

$$\begin{aligned} \Psi_+ &= \frac{a_\uparrow^+}{\sqrt{2}}\begin{bmatrix}1\\1\end{bmatrix} + \frac{a_\downarrow^+}{\sqrt{2}}\begin{bmatrix}1\\-1\end{bmatrix} \\ &= \frac{1}{\sqrt{2}}\begin{bmatrix}sin\zeta - cos\zeta e^{i\xi}\\ sin\zeta + cos\zeta e^{i\xi}\end{bmatrix}. \end{aligned} \tag{7.105}$$

The spin components in the lower and upper spin-split states are:

$$S_x^- = (\hbar/2)cos(2\zeta); \quad S_y^- = -(\hbar/2)sin(2\zeta)sin(\xi); \quad S_z^- = (\hbar/2)sin(2\zeta)cos(\xi) \tag{7.106}$$

and

$$S_x^+ = -(\hbar/2)cos(2\zeta); \quad S_y^+ = (\hbar/2)sin(2\zeta)sin(\xi); \quad S_z^+ = -(\hbar/2)sin(2\zeta)cos(\xi). \tag{7.107}$$

Thus, the spins in the upper and lower energy states are anti-parallel, as they should be.

One can see from the above that it is possible to "rotate" the ground state electron's spin polarization using the applied electric field in the y-direction which induces the Rashba interaction. By changing this electric field, we can change η and therefore D (see Equation (7.95)) and δ (see Equation (7.101)) and finally ζ. This causes the spin polarization vector to rotate. However, we cannot rotate the spin by arbitrary angle since δ can only be varied between 0 and ∞, so that ζ can be varied between 0 and $\pi/4$ radians only.

7.5 Why are the dispersion relations important?

After all is said and done, the reader may wonder what use do the dispersion relations have? We have led them through all this complicated algebra, enough to leave even the most patient and trained minds reeling, but do these dispersion relations have any practical use? The answer is an emphatic "yes". Optical properties of semiconductors, such as absorption and emission of polarized light, are determined by the spin resolved energy dispersion relations. Therefore, these dispersion relations are extremely important in opto-spintronics. They are also important in spin transport since spin current under a given set of circumstances may depend critically on the band structure. One can use the band structure properties to measure the spin-orbit interaction strengths, such the Rashba interaction strength η, using optical absorption and emission studies. Normally, these strengths are measured using transport experiments [6] but optical studies are always more convenient and "user-friendly" than transport studies, particularly in the case of nanostructures, since one does not have to make contacts to the structure to study its optical properties. Finally, many new spin devices have been proposed based on the peculiarities of the band structure. These include switching transistors [7] and ultra-sensitive magnetic field sensors [4]. Therefore, the dispersion relations are vital in more than one way.

The dispersion relations are also important in understanding certain intriguing phenomena. One of them is the *Intrinsic Spin Hall Effect* that is discussed next. The other is the *Spin Galvanic Effect* discussed in the next Chapter.

7.6 The three types of Hall Effect

Consider a solid structure subjected to a magnetic field of flux density B_z in the z-direction. An electric field is applied in the x-direction that causes an electron in the solid to accelerate in the -x-direction. However, scattering causes a frictional force, proportional to the electron velocity, which opposes the force due to the electric field. When the two forces balance, the electron reaches a constant velocity known as the 'drift velocity'. In the steady state, the electron drifts along the -x-direction with this constant velocity v_{drift}. The reader should be familiar with this model which is known as the *Drude model* of conduction.

The normal Hall Effect

The z-directed magnetic field exerts a Lorentz force $ev_{drift}B_z$ on the electron which pushes it in either the +y-direction or -y-direction depending on the direction of the magnetic field. If the sample has boundaries in the y-direction, then electrons will pile up at one edge (the edge that they are being pushed towards), causing a charge imbalance between this edge and the opposite edge. This charge imbalance will cause a y-directed electric field and an associated potential difference. The reader should recognize this as the celebrated *Hall Effect* that Edwin Hall had predicted in 1879 [8], and is routinely used today in semiconductor characterization to ascertain carrier concentration in a sample or its polarity (p-type or n-type). The sign of the Hall voltage, which is the potential difference between the edges, depends on the polarity of charge carriers (whether they are electrons or holes). Its magnitude is inversely proportional to the carrier concentration and directly proportional to the magnetic field.

The anomalous Hall Effect

In a ferromagnet, there can be an additional contribution to the Hall voltage that arises without any external magnetic field. Because of spin-dependent band structure or spin-dependent scattering events, electrons whose spins are polarized in the, say, +z-direction, are scattered to one edge of the sample and electrons whose spins are polarized in the -z-direction are scattered to the other edge. Since the material is a *ferromagnet*, there are majority and minority spins, meaning that there can be more electrons with +z-polarized spins than -z-polarized spins. In that case, more electrons are scattered towards one edge than the other, leading once again to a charge imbalance between these two edges and a resulting Hall voltage. This Hall voltage does not require a magnetic field. This is the so-called Anomalous Hall Effect, whose existence is often used as proof of ferromagnetism in materials.

The normal and anomalous Hall effects are explained in Fig. 7.9, which is adapted from ref. [9].

The Intrinsic Spin Hall Effect

This in an intriguing phenomenon that takes place in a two-dimensional electron gas (2-DEG) in ballistic transport, provided there is Rashba spin orbit interaction. When an electric field is applied in the plane of the 2-DEG, a dissipationless spin current (consisting of spins polarized perpendicular to the plane of the 2-DEG) flows in the plane of the 2-DEG in a direction perpendicular to a driving electric field. It is caused by the Rashba spin-orbit interaction [12, 13]. No external magnetic field is required. Because of this current, spins of one polarization (+z-polarized) pile up at one edge, and spins of the opposite polarization (-z-polarized) pile up at the other edge. However, because the material is not a ferromagnet, it has equal populations of electrons with +z-polarized and -z-polarized spin. As a result, no Hall voltage builds up between the two edges since no charge imbalance results from this

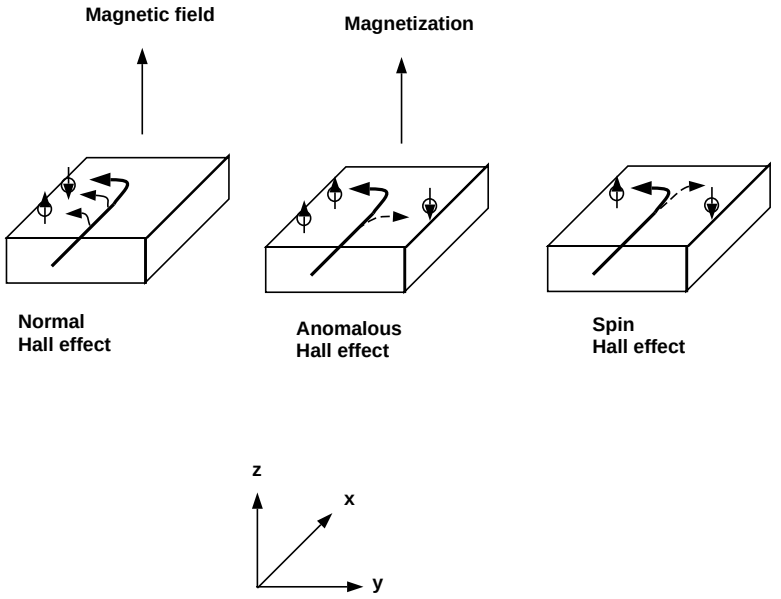


FIGURE 7.9
Pictorial depiction of the normal, anomalous and the spin Hall effect.

effect. However, a spin imbalance forms just like in the case of the anomalous Hall effect. This drives a spin current J_{sy} that flows between the two edges, in the y-direction. It is dissipationless since there is no electric field in the y-direction, so that $\vec{J}_{sy} \cdot \vec{E} = 0$ where $\vec{E}$ is the electric field in the 2-DEG.

Spin dependent scattering, or band structure peculiarities can also cause a spin imbalance between the two edges (in a paramagnet) and result in a spin current. That is known as the Extrinsic Spin Hall Effect [14], which has been experimentally demonstrated [15]. Here, we will discuss the Intrinsic Spin Hall Effect.

We will derive the Intrinsic Spin Hall Effect along the lines of ref. [12]. Consider a 2-DEG in the x-y plane (see Fig. 7.9) and a symmetry breaking electric field $\mathcal{E}_z$ is applied in the z-direction to induce the Rashba spin-orbit interaction in the electron gas. The Hamiltonian describing this system is

$$\begin{aligned} H &= \frac{|\vec{p}|^2}{2m^*} + V(\vec{r}) - \frac{a_{46}}{\hbar}\mathcal{E}_z\hat{z}\cdot[\vec{\sigma}\times\vec{p}] \\ &= \begin{bmatrix} \frac{p^2}{2m^*} + V(\vec{r}) & \frac{\eta}{\hbar}(p_y + ip_x) \\ \frac{\eta}{\hbar}(p_y - ip_x) & \frac{p^2}{2m^*} + V(\vec{r}) \end{bmatrix} , \end{aligned} \tag{7.108}$$

where $V(\vec{r})$ is the potential energy, $\vec{p} = p_x\hat{x} + p_y\hat{y}$ and $p = \sqrt{p_x^2 + p_y^2}$.

The Rashba spin-orbit interaction acts like an effective magnetic field B_{Rashba}. We know that spin-orbit interaction gives rise to a magnetic field in the reference frame of a moving electron moving in an electric field. This magnetic field is (see Equation (6.2))

$$\vec{B}_{Rashba} = \frac{\vec{E}\times\vec{v}}{2c^2} , \tag{7.109}$$

where $\vec{v}$ is the electron velocity, $\vec{E}$ is the electric field and c is the speed of light in vacuum. Here we have ignored the relativistic factor since we are dealing with a non-relativistic electron in a solid. Recently, the effective magnetic fields associated with Rashba and Dresselhaus spin-orbit interactions have been experimentally measured in a two-dimensional electron gas [10]. Note that according to the above equation, the effective field depends on the velocity $\vec{v}$ and *not* the momentum $\vec{p}$ or $\hbar\vec{k}$. It *makes a difference*, since in the presence of spin-orbit interaction, $\vec{v}$ and $\vec{p}$ (or $\vec{k}$) are *not* proportional to each other, as we have already seen (recall Section 7.1.1.1 and Equation (7.32)). We can show this more directly by considering the fact that the Hamiltonian of an electron in a solid, in the presence of spin-orbit interaction, is given by

$$H = \frac{\hbar^2k^2}{2m^*} + V(\vec{r}) + \frac{\hbar^2}{4(m^*)^2c^2}\left[\frac{\partial V(\vec{r})}{\partial\vec{r}}\right]\cdot\left[\vec{\sigma}\times\vec{k}\right] . \tag{7.110}$$

Hamiltonian mechanics for canonically conjugated variables [momentum ($\vec{p}$)

and position ($\vec{r}$)] dictate

$$\vec{v} = \frac{d\vec{r}}{dt} = \frac{\partial H}{\partial \vec{p}} = \frac{\partial H}{\partial(\hbar\vec{k})} = \frac{\hbar\vec{k}}{m^*} + \frac{\hbar^2}{4(m^*)^2c^2}\left[\frac{\partial V(\vec{r})}{\partial \vec{r}} \times \vec{\sigma}\right]. \tag{7.111}$$

Clearly, because of the second term on the right hand side, which is entirely due to spin-orbit interaction, $\vec{v}$ is not proportional to $\vec{k}$. They are related as $\vec{v} = A\vec{k} + B$, where A and B are constants. This is not a bandstructure issue; it is true even in a parabolic band. Therefore, B_{Rashba} cannot be simultaneously proportional to both v and k. It should be one or the other. Unfortunately, this issue is not always appreciated. One often finds in the literature that B_{Rashba} is assumed proportional to k [11], and sometimes it is assumed proportional to v. Obviously, both cannot be correct in every situation. Here, we will take the first steps towards resolving this issue.

One can appeal to the Ehrenfest Theorem (which we saw in Chapter 4) to deduce whether B_{Rashba} is proportional to v or k. According to this theorem (see Chapter 15 for the exact statement of this theorem), we get

$$\frac{d\langle\vec{S}\rangle}{dt} = \frac{1}{i\hbar}\langle[H, \vec{S}]\rangle = \frac{1}{i\hbar}\langle[H, (\hbar/2)\vec{\sigma}]\rangle, \tag{7.112}$$

where the angular brakets denote the expectation values and the square bracket denotes the commutator. Here H is the Hamiltonian in Equation (7.108). However Equation (4.4) tells us that the left hand side of the above equation is $(g\mu_B\vec{B}_{Rashba}/\hbar) \times \langle\vec{S}\rangle$. Therefore, we obtain that

$$\frac{g\mu_B\vec{B}_{Rashba}}{\hbar} \times \langle\vec{S}\rangle = \frac{1}{i\hbar}\langle[H, \vec{S}]\rangle = \frac{1}{i\hbar}\left\langle\left[-\frac{a_{46}}{\hbar}\mathcal{E}_z\hat{z}\cdot[\vec{\sigma}\times\vec{p}], (\hbar/2)\vec{\sigma}\right]\right\rangle. \tag{7.113}$$

The reader can easily show from the above equation that B_{Rashba} is proportional to p or k, and not v[§]. Thereafter, it is easy to show that B_{Rashba} can be obtained by equating the Zeeman splitting energy caused by B_{Rashba} to the spin-splitting energy $2\eta k$ which would yield that

$$|B_{Rashba}| = \frac{2\eta k}{g\mu_B} = \left|\frac{2a_{46}\mathcal{E}_z k}{g\mu_B}\right| \tag{7.114}$$

However, this result presents an immediate conundrum. First, note from Fig. 7.4(b)[¶] that an electron with a given energy E does *not* have a unique wavevector k in the presence of spin-orbit interaction. The wavevector is *spin-dependent* since the electron will have two different wavevectors in the two spin-split bands. The wavevector will be $k_1(E)$ in one spin band and

[§]The authors are grateful to Prof. Yuli Lyanda-Geller for this discussion.

[¶]The quantum wire can be thought of as a 2-DEG with edges.

$k_2(E)$ in the other. Therefore, if $|B_{Rashba}|$ were proportional to k according to Equation (7.114), then it would be proportional to k_1 in one band and k_2 in the other. Thus, it will have two different magnitudes for two different spins. In other words, the spin-orbit magnetic field that an electron experiences *depends on what its spin orientation is.*

A *spin-dependent* B_{Rashba} is fine as long as we remember that it is different in the two spin-split bands and account for that fact explicitly in any analysis. However, there are situations when a spin-dependent field becomes inconvenient. This happens if spins are changing continuously with time which then necessitates a time dependent B_{Rashba}. If the spins are changing randomly because of spin flip scattering, then B_{Rashba} changes randomly with time. If the spins are precessing coherently in time, B_{Rashba} still changes continuously, albeit coherently. It is therefore much more convenient to deal with a *spin-independent* B_{Rashba} since that could be time-invariant even if the spins are time-varying. In the next chapter, we will show that such a B_{Rashba} or $B_{Dresselhaus}$ can be formulated and will be proportional to v and not k. The spin-independent fields are very convenient tools to explain the workings of the Spin Field Effect Transistor, or the absence of D'yakonov-Perel' spin relaxation in strictly single channeled transport.

In the discussion of Intrinsic Spin Hall Effect, we can however work with a spin-dependent B_{Rashba} since we will account for effects arising from the two spin-split bands separately. Therefore, we will use Equation (7.114). The reader is warned that the magnetic field in Equation (7.114) is always *spin-dependent* and using this equation requires particular care; we must account for effects from the two spin-split bands separately.

Before we proceed further, we need to ascertain the direction of B_{Rashba} which is not given by Equation (7.114). Equation (7.31) and Equation (7.111) both tell us that v and k are collinear, i.e., they have the same direction. Next, we appeal to Equation (6.2) which tells us that the direction of B_{Rashba} (in either spin-split band) is mutually perpendicular to the direction of the effective electric field that causes the Rashba interaction and the velocity (or wavevector). Therefore, for our 2-DEG, we can write the *spin-dependent* B_{Rashba} as

$$\begin{aligned} B_{Rashba} &= -\frac{2a_{46}}{g\mu_B}\vec{\mathcal{E}} \times \vec{k} = \frac{2\eta}{g\mu_B}\hat{z} \times \vec{k} \\ &= \Xi\left[-k_y\hat{x} + k_x\hat{y}\right] \ , \end{aligned} \tag{7.115}$$

where $\Xi = \frac{2\eta}{g\mu_B}$. Note that $\vec{k} \cdot \vec{B}_{Rashba} = 0$.

Next, using Equation (4.4), we get

$$\begin{aligned} \frac{d\vec{S}}{dt} &= \vec{\Omega} \times \vec{S} \\ &= \frac{g\mu_B \vec{B}_{Rashba}}{\hbar} \times \vec{S} \end{aligned}$$

$$= det \begin{bmatrix} \hat{x} & \hat{y} & \hat{z} \\ \frac{g\mu_B}{\hbar} B_x & \frac{g\mu_B}{\hbar} B_y & 0 \\ S_x & S_y & S_z \end{bmatrix}$$

$$= \frac{2\eta}{\hbar} \left[k_x S_z \hat{x} + k_y S_z \hat{y} - (k_y S_y + k_x S_x) \hat{z} \right] \ , \qquad (7.116)$$

where B_x and B_y are the x- and y-components of B_{Rashba}.

Therefore,

$$\frac{dS_x}{dt} = \frac{g\mu_B}{\hbar} B_y S_z = \frac{2\eta}{\hbar} k_x S_z$$

$$\frac{dS_y}{dt} = -\frac{g\mu_B}{\hbar} B_x S_z = \frac{2\eta}{\hbar} k_y S_z$$

$$\frac{dS_z}{dt} = \frac{g\mu_B}{\hbar} (B_x S_y - B_y S_x) = -\frac{2\eta}{\hbar} (k_y S_y + k_x S_x). \qquad (7.117)$$

Consider now the situation when an electric field E_x is applied in the positive x-direction at time $t = 0$. Let us focus on an electron with initial velocity $\vec{v}_0$ and wavevector $\vec{k}_0$ at time t=0 and define a new set of coordinate axes (the 'primed' coordinates) such that the y'-axis coincides with $\vec{v}_0$ or $\vec{k}_0$ (see Fig. 7.10). It should be obvious that Equations (7.115 - 7.117) will continue to hold if we replace the unprimed coordinates everywhere with the primed coordinates since the $x - y$ system in these equations is arbitrary. Let the y'-axis subtend an angle ϕ with the y-axis.

At time t =0, $B_{Rashba}(t = 0)$ is along the x' direction, since it is perpendicular to $\vec{k}_0$ and $\hat{z}$. Since $B_{x'} \neq 0$ and $B_{y'} = 0$, $S_{x'}(0)$ = -$\hbar$/2 in the lower Rashba spin-split band, and +$\hbar$/2 in the upper spin-split band, while $S_{y'}(0)$ = $S_z(0) = 0$ in either band. The application of the electric field along the x-direction will change B_{Rashba} since it will accelerate the electron and change k_x (see Equation (7.115)). We will assume that the electric field E_x is so weak that B_{Rashba} changes slowly in time, allowing the spin to follow adiabatically and align itself parallel or anti-parallel to B_{Rashba}, at every instant of time.

Immediately after the application of the electric field, i.e., at time $t = 0+ = \tau$, the change in the magnitude of B_{Rashba} can be ignored, but its direction will have changed slightly. Let us assume that it has rotated by a small angle $\delta(\tau)$ from its initial direction at time $t = 0$. Then, the change in $B_{x'}$ is $B_{x'}(cos\delta - 1) \approx -B_{x'}\delta^2$, while the change in $B_{y'}$ is $B_{x'} sin\delta \approx B_{x'}\delta$. The ratio of the two changes is -δ, which is a small quantity. Therefore, the former change is much smaller than the latter. Consequently, we can postulate that at $t = 0+ = \tau$

$$\frac{dB_{x'}(t)}{dt} \approx 0$$

$$\frac{dB_{y'}(t)}{dt} \neq 0$$

$$S_{x'}(t) \approx \pm\tfrac{\hbar}{2} \ , \qquad (7.118)$$

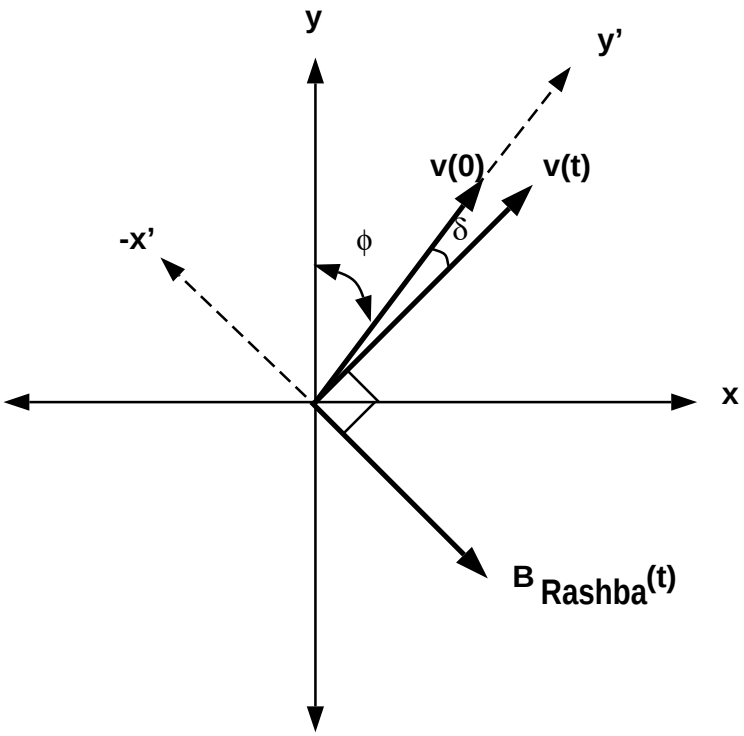


FIGURE 7.10
The coordinate axes in the two-dimensional electron gas and the velocity vector, as well as the vector representing B_{Rashba}.

where the '+' sign is for the upper Rashba spin-split band and the '-' sign for the lower Rashba spin-split band.

At time $t = 0+ = \tau$, B_{Rashba} develops a small y'-component. Since the spin follows B_{Rashba} adiabatically, we infer that

$$\frac{dS_{y'}}{dt} \neq 0. \tag{7.119}$$

We will argue that the change is linear in time over a small duration τ so that

$$\frac{d^2S_{y'}}{dt^2} = 0. \tag{7.120}$$

From Equations (7.117) - (7.120), we get that at time $t = \tau$

$$0 = \frac{d^2S_{y'}}{dt^2} = -\frac{g\mu_B}{\hbar}\left[\frac{dB_{x'}(t)}{dt}S_z + \frac{dS_z}{dt}B_{x'}(t)\right] \approx -\frac{g\mu_B}{\hbar}\frac{dS_z}{dt}B_{x'}. \tag{7.121}$$

Using Equation (7.117) in the above equation, we get

$$\begin{aligned} 0 &\approx B_{x'}(\tau)S_{y'}(\tau) - B_{y'}(\tau)S_{x'}(\tau) \\ &\approx B_{x'}(\tau)S_{y'}(\tau) \mp B_{y'}(\tau)(\hbar/2)\ , \end{aligned} \tag{7.122}$$

which yields

$$S_{y'}(\tau) = \pm\frac{B_{y'}(\tau)}{B_{x'}(\tau)}\frac{\hbar}{2}\ , \tag{7.123}$$

where the '+' sign is for the upper Rashba band and the '-' sign is for the lower Rashba band.

From Equation (7.117), we also get

$$\frac{dS_{y'}}{dt} = -\frac{g\mu_B}{\hbar}B_{x'}(t)S_z(t)\ . \tag{7.124}$$

Substitution of Equation (7.123) in Equation (7.124) yields that at $t = \tau$

$$\begin{aligned} S_z(t) &= \mp\frac{\hbar}{g\mu_B}\frac{1}{B_{x'}(t)}\frac{d}{dt}\left(\frac{B_{y'}(t)}{B_{x'}(t)}\right)\frac{\hbar}{2} \\ &= \mp\frac{\hbar}{g\mu_B}\frac{1}{B_{x'}(t)}\left[\frac{B_{x'}(t)(dB_{y'}(t)/dt) - B_{y'}(t)(dB_{x'}(t)/dt)}{B^2_{x'}(t)}\right]\frac{\hbar}{2} \\ &= \mp\frac{\hbar}{g\mu_B}\frac{1}{B^2_{x'}(t)}\frac{dB_{y'}(t)}{dt}\frac{\hbar}{2}\ , \end{aligned} \tag{7.125}$$

where we have used Equation (7.118) to arrive at the last equality above.

Preceding Equation (7.118), we stated that the change in $B_{y'}$ is

$$B_{y'}(\tau) - B_{y'}(0) = B_{y'}(\tau) \approx B_{x'}(\tau)\delta(\tau). \tag{7.126}$$

Substituting this in Equation (7.125), we obtain that at $t = \tau$

$$S_z(\tau) = \mp\frac{\hbar^2}{2g\mu_B}\frac{1}{B_{x'}^2(t)}\left[\frac{dB_{x'}(t)}{dt}\delta(t) + B_{x'}(t)\frac{d\delta(t)}{dt}\right] \approx \mp\frac{\hbar^2}{2g\mu_B B_{x'}(\tau)}\frac{d\delta(t)}{dt}\ , \tag{7.127}$$

where we have used the fact that $dB_{x'}/dt \approx 0$.

Referring to Fig. 7.10,

$$sin(\phi + \delta(t)) = \frac{v_x(t)}{v(t)} = \frac{k_x(t)}{k(t)}, \tag{7.128}$$

where the last equality above can be easily proved using Equation (7.31). Therefore,

$$\begin{aligned}
\frac{d\delta(t)}{dt} &= \frac{d(\phi + \delta(t))}{dt} = \frac{d}{dt}\left[sin^{-1}\left(\frac{k_x(t)}{k(t)}\right)\right] \\
&= \frac{1}{\sqrt{1 - k_x^2(t)/k^2(t)}}\frac{d}{dt}\left(\frac{k_x(t)}{k(t)}\right) \\
&= \frac{1}{k_y(0)/k(t)}\frac{k(t)(dk_x(t)/dt) - k_x(t)(dk(t)/dt)}{k^2(t)} \\
&= \frac{1}{k_y(0)/k(t)}\frac{k(t)(dk_x(t)/dt) - (k_x^2(t)/k(t))(dk_x(t)/dt)}{k^2(t)} \\
&= \frac{k(t)}{k_y(0)}\frac{k_y^2(0)}{k^3(t)}\frac{dk_x(t)}{dt} \\
&= \frac{k_y(0)}{k^2(t)}\frac{dk_x(t)}{dt},
\end{aligned} \tag{7.129}$$

where we have used the fact that $k_y(t) = k_y(0)$ because there is no electric field in the y-direction.

Substituting Equation (7.129) in Equation (7.127), we obtain

$$S_z(\tau) = \mp\frac{\hbar^2}{2g\mu_B B_{x'}(\tau)}\frac{k_y(0)}{k^2(\tau)}\frac{dk_x}{dt}. \tag{7.130}$$

Newton's law in a crystal dictates that

$$\frac{d(\hbar k_x)}{dt} = -eE_x, \tag{7.131}$$

where we have neglected forces due to scattering, assuming ballistic transport. The above equation is valid even in the presence of spin-orbit interaction (see Problem 7.6).

Substituting the above result in Equation (7.130), we obtain

$$S_z(\tau) = \pm\frac{e\hbar}{2g\mu_B B_{x'}(\tau)}\frac{k_y(0)}{k^2(\tau)}E_x. \tag{7.132}$$

Now, from Fig. 7.10,

$$B_{x'}^2 = B_{Rashba}^2 = B_x^2 + B_y^2, \tag{7.133}$$

where B_x is the x-component and B_y is the y-component of $B_{x'}$ or B_{Rashba}. Using Equation (7.115) in the above equation, we get

$$B_{x'}^2 = \Xi^2(k_x^2 + k_y^2) = \Xi^2 k^2(t). \tag{7.134}$$

Using this relation in Equation (7.132) yields

$$S_z(\tau) = \pm\frac{e\hbar}{4\eta}\frac{k_y(0)}{k^3(\tau)}E_x \ , \tag{7.135}$$

where the '-' sign is for the lower spin band and '+' sign for the upper.

Let us now calculate the y-directed spin (angular momentum) current density $J_{sy}(t)$ at temperature $T \to 0$.

The spin current density operator associated with z-polarized spin is $j_{s_z} = (1/2)\{\sigma_z, \vec{v}\}$, where the curly brackets denote the anti-commutator. Its expected value is given by (see [12]) $< \psi^+|j_{s_z}|\psi^- >$, where $|\psi^\pm >$ are the eigenstates in the upper and lower spin bands. The reader can easily show that the eigenstates are

$$|\psi^\pm >= \frac{1}{\sqrt{2}}\begin{pmatrix} \pm e^{i\theta} \\ 1 \end{pmatrix} e^{i\vec{k}\cdot\vec{r}}, \tag{7.136}$$

where $\theta = arctan(k_x/k_y)$.

The operator v_y is:

$$v_y = \frac{\partial H}{\partial p_y} = \begin{bmatrix} \frac{p_y}{m^*} & \frac{\eta}{\hbar} \\ \frac{\eta}{\hbar} & \frac{p_y}{m^*} \end{bmatrix}. \tag{7.137}$$

Therefore, the operator for the y-component of the z-polarized spin current density is

$$j_{s_z}^y = \frac{1}{2}\{\sigma_z, v_y\} = \begin{bmatrix} \frac{p_y}{m^*} & 0 \\ 0 & \frac{p_y}{m^*} \end{bmatrix}. \tag{7.138}$$

Consequently, the contribution to the y-component of the z-polarized spin current density from a wavevector state $\vec{k}$ in either spin-split band is $J_{sy}(k(\tau))$ $= S_z^\pm(k(\tau)) < \psi^+|j_{s_z}^y|\psi^- > = S_z^\pm(k(\tau))(\hbar k_y(0)/m^*)$.

Therefore, the y-directed spin current density due to z-polarized spin at time $t = \tau$ is

$$J_{sy}(\tau) = \frac{1}{4\pi^2}\int_0^{2\pi}\int_0^{k_F} d\phi k(\tau)dk(\tau)\frac{\hbar k_y(0)}{m^*}S_z^\pm(k(\tau)) \ , \tag{7.139}$$

where k_F is the Fermi wavevector, which is the wavevector corresponding to the Fermi energy (see Fig. 7.11). The spin current density in each band is:

$$J_{sy1}(\tau) = \frac{e\hbar^2}{16\pi^2 m^*\eta}E_x\int_0^{2\pi}\int_0^{k_{F1}} d\phi k(\tau)dk(\tau)\frac{k_y^2(0)}{k^3(\tau)} \quad (upper\ band)$$

$$J_{sy2}(\tau) = -\frac{e\hbar^2}{16\pi^2 m^*\eta} E_x \int_0^{2\pi} \int_0^{k_{F2}} d\phi k(\tau) dk(\tau) \frac{k_y^2(0)}{k^3(\tau)} \quad (lower\ band). \tag{7.140}$$

Since $k_y(0) \approx k(\tau) cos\phi$ (see Fig. 7.10), we obtain

$$\begin{aligned} J_{sy1}(\tau) &= \frac{e\hbar^2}{16\pi^2 m^*\eta} E_x \int_0^{2\pi} \int_0^{k_{F1}} d\phi dk(\tau) cos^2\phi \\ &= \frac{e\hbar^2}{16\pi m^*\eta} E_x k_{F1} \quad (upper\ band) \\ J_{sy2}(\tau) &= -\frac{e\hbar^2}{16\pi^2 m^*\eta} E_x \int_0^{2\pi} \int_0^{k_{F2}} d\phi dk(\tau) cos^2\phi \\ &= -\frac{e\hbar^2}{16\pi m^*\eta} E_x k_{F2} \quad (lower\ band). \end{aligned} \tag{7.141}$$

Consequently, if both Rashba spin-split bands are occupied, then the total spin current is

$$J_{sy}(\tau) = J_{sy1}(\tau) + J_{sy2}(\tau) = -\frac{e}{8\pi} E_x, \tag{7.142}$$

where we have used the fact that $k_{F2} - k_{F1} = 2m^*\eta/\hbar^2$.

The spin Hall conductivity is

$$\sigma_{SH} = \frac{J_{sy}}{-E_x} = \frac{e}{8\pi}. \tag{7.143}$$

Note that the spin Hall conductivity depends only on the universal constant e, which is the charge of the electron. Hence the name "universal" intrinsic Spin Hall Effect. Note also that we accounted for contributions from the two spin-split bands separately, which allowed us to work with the spin-dependent B_{Rashba}.

7.6.1 Quantum spin-Hall Effect

The integer Quantum Hall Effect [16] was discovered more than 25 years ago and resulted in the 1985 Nobel Prize in physics. This effect makes the Hall resistance of a sample (which is the Hall voltage divided by the current flowing through the sample) quantized. It occurs in a two-dimensional electron gas at high magnetic fields and results from the formation of conducting one dimensional channels that develop at the edges of the sample (hence the name "edge states"). Each of these edge channels transmits with unit probability (regardless of scattering - which is the surprising result) and contributes a quantized conductance of $2e^2/h$[‖]. The Hall resistance is also quantized and

[‖] We will ignore the fact that the magnetic field can lift the spin degeneracy of the edge states because of the Zeeman effect. The g-factors in most materials are small enough that

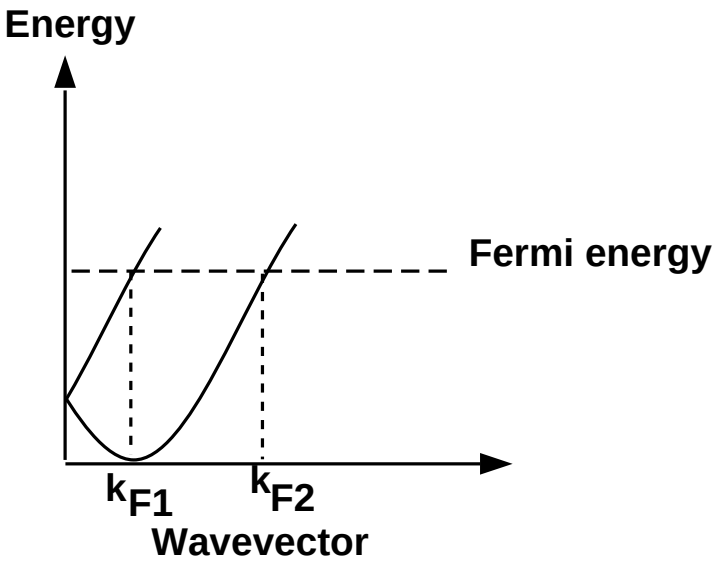


FIGURE 7.11
Energy dispersion relation in a two-dimensional electron gas with Rashba spin-orbit interaction. The two-dimensional wavevector in the x-y plane is denoted by "Wavevector".

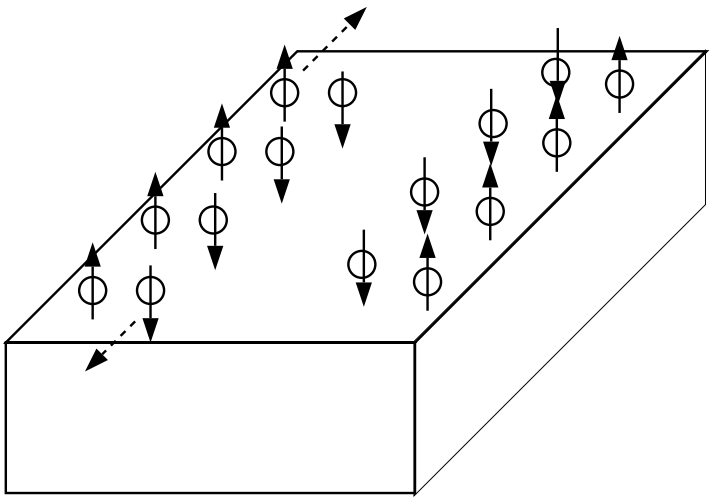


FIGURE 7.12
Depiction of the Quantum spin Hall Effect where edge channels consist of spin polarized carriers. In any given edge, channels with opposite spin polarization travel in opposite directions as shown by the broken arrows.

given by

$$R_{quant-Hall} = \frac{h}{2e^2 N}, \tag{7.144}$$

where N is the number of edge channels.

Each edge channel conducts a dissipationless current in one direction. They normally form only at high magnetic fields, but recently there has been theoretical predictions that a different kind of edge channel can form even without a magnetic field [17, 18, 19]. They will form in insulators with suitable electronic structures. Carriers with opposite spins will move in opposite directions on a given edge as shown in Fig. 7.12. This is the *Quantum spin Hall Effect*.

Bernevig, et al. [20] recently predicted that a quantum well of HgTe will exhibit the Quantum spin Hall Effect because of the peculiarities of its electronic structure. This has now been confirmed experimentally culminating in the demonstration of the Quantum spin Hall Effect [21]. Research in this field is just beginning to take root and exciting discoveries might be around the

they cause negligible Zeeman splitting at the magnetic fields where the Quantum Hall Effect is observed. Hence, we can assume that the edge states are spin-degenerate.

corner. Reference [21] appeared in the literature at the time when the galley proof of this book became available. This is a development to watch since fundamental discoveries may ensue.

7.7 Problems

- **Problem 7.1**

  Derive the second equality in Equation (7.41).

  Solution

  We have

$$\left[\frac{(p_x + eBz)^2 + p_z^2}{2m^*}\right] e^{ik_x x}\zeta_n(z) = \varepsilon_n e^{ik_x x}\zeta_n(z). \tag{7.145}$$

  This can be re-written as

$$\left[\frac{p_x^2}{2m^*} + \frac{p_x eBz}{m^*} + \frac{e^2B^2z^2}{2m^*} + \frac{p_z^2}{2m^*}\right] e^{ik_x x}\zeta_n(z) = \varepsilon_n e^{ik_x x}\zeta_n(z) \tag{7.146}$$

  since the operators p_x and z commute. Next, multiplying by the "bra" $e^{-ik_x x}$ and integrating over the coordinate x, we get

$$\left[\frac{\hbar^2 k_x^2}{2m^*} + \frac{\hbar k_x eBz}{2m^*} + \frac{e^2B^2z^2}{2m^*} + \frac{p_z^2}{2m^*}\right] \zeta_n(z) = \varepsilon_n\zeta_n(z). \tag{7.147}$$

  Using the definition of z_0, we can re-write the above equation as

$$\frac{e^2B^2z_0^2}{2m^*} + \frac{e^2B^2 2zz_0}{2m^*} + \frac{e^2B^2z^2}{2m^*} + \frac{p_z^2}{2m^*} = \varepsilon_n\zeta_n(z). \tag{7.148}$$

  Finally, using the definition of cyclotron frequency ω_c, we get from the above

$$\left[\frac{p_z^2}{2m^*} + \frac{m^*}{2}\omega_c^2\left(z + z_0\right)^2\right] \zeta_n(z) = \varepsilon_n\zeta_n(z). \tag{7.149}$$

- **Problem 7.2**

  Show that if the thickness of a 2-DEG is vanishingly small, then the Dresselhaus interaction term can be written as in Equation (7.6).

  Solution

  Because there are no extended states in the direction of confinement in a 2-DEG (y-direction in our case), the expected value of the y-component of momentum $< p_y > = 0$. This makes $< \kappa_y > = 0$ (see Equation

(7.3)) where the spatial averaging is carried out over the y-coordinate only. If the 2-DEG is vanishingly thin, then $< p_y^2 > \rightarrow \infty$. Therefore, $< \kappa_x > (p_x + eA_x) \times constant$, and $< \kappa_z > =$ -$(p_z + eA_z) \times constant$. Defining ν such that the *constant* is equal to -$\nu/\hbar$ allows us to write the Dresselhaus interaction in the form given in Equation (7.6).

- **Problem 7.3**

  Consider the case of a quantum wire with spin-orbit interaction and a magnetic field directed along the axis of the wire. Show that the dispersion relation is symmetric about the energy-axis ($k_x = 0$) only if the Dresselhaus interaction is absent. Show that when the magnetic field is along the direction of the electric field causing the Rashba interaction, the dispersion relation is always symmetric about the energy-axis and when the magnetic field is along the direction mutually perpendicular to the axis of the quantum wire and the electric field causing the Rashba interaction, the dispersion relation is symmetric about the energy-axis only if $\alpha_3\Delta = \eta\beta$.

  Solution

  Case I: When the magnetic field is along the wire axis.

  In this case, if the Dresselhaus interaction is absent, then $\nu = 0$ and the dispersion relation in Equation (7.49) simplifies to

$$E_{\pm}^{(1)} = \frac{\hbar^2 k_x^2}{2m^*} + E_0 \pm \sqrt{(\eta k_x)^2 + \beta^2}. \tag{7.150}$$

  It is obvious from the above that $E_{\pm}^{(1)}(k_x) = E_{\pm}^{(1)}(-k_x)$ which proves that the dispersion relation is symmetric about the energy-axis.

  Case II: When the magnetic field is along the direction of the electric field causing the Rashba interaction.

  The dispersion relation is

$$E_{\pm}^{(2)} = \frac{\hbar^2 k_x^2}{2m^*} + E_0' \pm \sqrt{(\eta^2 + \alpha_2^2)\, k_x^2 + \beta^2}. \tag{7.151}$$

  In this case, $E_{\pm}^{(2)}(k_x) = E_{\pm}^{(2)}(-k_x)$, always.

  Case III: When the magnetic field is along the direction mutually perpendicular to the axis of the quantum wire and the electric field causing the Rashba interaction, and $\alpha_3\Delta = \eta\beta$.

  The dispersion relation is

$$E_{\pm}^{(3)} = \frac{\hbar^2 k_x^2}{2m^*} + E_0' \pm \sqrt{(\eta^2 + \alpha_3^2)\, k_x^2 + +\Delta^2 + \beta^2}. \tag{7.152}$$

  Here too, $E_{\pm}^{(2)}(k_x) = E_{\pm}^{(2)}(-k_x)$, which makes the dispersion relations symmetric about the energy-axis.

- **Problem 7.4**

Consider the case when a magnetic field is directed along the axis of a quantum wire and the Dresselhaus interaction is absent. Show that the minima of the upper spin resolved subband occurs at $k_x = 0$, but the minima of the lower subband occur at $k_x = \pm\sqrt{\frac{m^{*2}\eta^2}{\hbar^4} - \frac{\beta^2}{\eta^2}}$. The lower subband also has a local maximum at $k_x = 0$. This gives rise to the "camel-back" feature in the lower subband seen in Fig. 7.4. Show that the "camel-back" feature disappears if the magnetic flux density $B > B_c$ where $B_c = \frac{2m^*\eta^2}{\hbar^2 g\mu_B}$. Show that the same is true in the case when the magnetic field is along the direction of the electric field causing the Rashba interaction. In the latter case, all of the above is true even if there is a Dresselhaus interaction. The value of the B_c then is given by $B_c = \frac{2m^*(\eta^2+\alpha_2^2)}{\hbar^2 g\mu_B}$.

Solution

To find the extrema, we differentiate the energy with respect to k_x, set the result equal to zero, and solve for k_x.

In this case,

$$\frac{1}{\hbar}\frac{\partial E_\pm}{\partial k_x} = \frac{\hbar k_x}{m^*} \pm \frac{\eta^2 k_x}{\hbar\sqrt{(\eta k_x)^2 + \beta^2}}. \tag{7.153}$$

Setting this equal to zero and solving for k_x, we get the following:

Upper band: In the case of the + sign, there is only one solution for k_x which is $k_x = 0$. This shows that the upper subband has a minimum at $k_x = 0$.

Lower band: In this case, there are two solutions for k_x. One is $k_x = 0$, which corresponds to a local maximum. The other solution is

$$k_x^2 = \frac{m^{*2}\eta^2}{\hbar^4} - \frac{\beta^2}{\eta^2}. \tag{7.154}$$

We note first that there is no *real* solution for k_x if $\beta > \eta^2 m^*/\hbar^2$, or $B > \frac{2m^*\eta^2}{\hbar^2 g\mu_B}$. Therefore, there are no other extrema if $B > B_c$ where $B_c = \frac{2m^*\eta^2}{\hbar^2 g\mu_B}$. In that case, there will be no "camel-back" feature. If, on the other hand, $B < B_c$, then from Equation(7.154), we get that there are minima located at

$$k_x = \pm\sqrt{\frac{m^{*2}\eta^2}{\hbar^4} - \frac{\beta^2}{\eta^2}}. \tag{7.155}$$

The reader can easily extend this analysis to show that in the case when the magnetic field is along the direction of the electric field inducing

the Rashba effect, the minima occur at $k_x = \pm\sqrt{\frac{m^*}{\hbar^2} - \frac{\beta^2}{\eta^2+\alpha_2^2}}$ and $B_c = \frac{2m^*(\eta^2+\alpha_2^2)}{\hbar^2 g\mu_B}$.

- **Problem 7.5**

  Consider a quantum dot in which there is no Dresselhaus interaction, but there is a Rashba interaction due to a symmetry breaking electric field in the y-direction. There is also a magnetic field in the x-direction. In this case, show that the spin of an electron in the ground state lies in the x-y plane, i.e., $S_z = 0$.

  Solution

  If only Rashba interaction is present, then from Equation (7.101), we see that $\delta = D$. However, since D is purely imaginary, we get that $\delta = i|\delta|$, so that the angle $\xi = \pi/2$. Using this result in Equation (7.106), we immediately find that $S_z = 0$. Note that when the Rashba interaction strength is extremely strong so that $|\delta| >> g\mu_B B$, then $\zeta = \pi/4$ and we have $S_x^- = S_z^- = 0$, and S_y^- = -1, while $S_x^+ = S_z^+ = 0$, and $S_y^+ = 1$. In that case, the ground state spin points against the electric field inducing the Rashba effect and the excited state spin points in the direction of the electric field. Of course, when the Rashba interaction strength is very weak so that $|\delta| << g\mu_B B$, then $S_y^\pm = S_z^\pm = 0$, and $S_x^\pm = \mp 1$. In that case, the ground state spin points along the magnetic field and the excited state spin points against the magnetic field, as we expect.

- **Problem 7.6**

  Show that Equation (7.131) is correct in the presence of spin-orbit interaction. Show also that

$$\frac{dv_x}{dt} = -\frac{eE_x}{m^*} \tag{7.156}$$

  even though $\hbar k_x \neq m^* v_x$.

  Hint: Use Ehrenfest Theorem.

  Show that if you used Equations (7.31) and (7.131) to derive dv_x/dt, you would not have obtained the same result as above. In fact, you would have obtained a wrong result. This is an example of **pitfalls** that lurk around when spin-orbit interaction is present.

- **Problem 7.7**

  Show that if the *spin-dependent* B_{Rashba} were proportional to velocity v instead of wavevector k, then the Intrinsic Spin Hall Effect would *not* be universal and the spin Hall conductivity would be

$$\sigma_{SH} = \frac{e}{8\pi}\left[1 + ln\left(\frac{\hbar v_F}{\eta}\right)\right], \tag{7.157}$$

where v_F is the Fermi velocity. You will end up with an improper integral $\int_{-\eta/\hbar}^{\eta/\hbar} \frac{1}{v(\tau)} dv(\tau)$. This has a singularity at $v(\tau) = 0$, but since the integrand is an odd function of the integration variable, this integral is exactly zero.

- **Problem 7.8**

  Since the B_{Rashba} that we used in the derivation of the universal Intrinsic Spin Hall Effect is spin-dependent, we should have accounted for two different B_{Rashba}-s (i.e., B^{+}_{Rashba} and B^{-}_{Rashba}). Repeat the derivation using two different B_{Rashba}-s for the two spin-split bands and show that the same result emerges.

7.8 References

[1] M. Cahay and S. Bandyopadhyay, "Phase coherent quantum mechanical spin transport in a weakly disordered quasi one-dimensional channel", Phys. Rev. B, **69**, 045303 (2004).

[2] U. Zülicke and M. Governale, "Spin accumulation in quantum wire with strong Rashba spin orbit coupling", Phys. Rev. B, **66**, 073311 (2002).

[3] S. Pramanik, S. Bandyopadhyay and M. Cahay, Energy dispersion relations of spin-split subbands in a quantum wire and electrostatic modulation of carrier spin polarization, Phys. Rev. B., **76**, 155325 (2007).

[4] S. Bandyopadhyay and M. Cahay, "Proposal for a spintronic femto-Tesla magnetic field sensor", Physica E, **27**, 98 (2005).

[5] S. Pramanik, S. Bandyopadhyay and M. Cahay, Phys. Rev. B, **76**, 155325 (2007)..

[6] B. Das, D. C. Miller, S. Datta, R. Reifenberger, W. P. Hong, P. K. Bhattacharyya, J. Singh and M. Jaffe, "Zero field spin splitting in a two-dimensional electron gas", Phys. Rev. B, **39**, 1411 (1989).

[7] J. Schliemann, J. C. Egues and D. Loss, "Non ballistic spin field effect transistor", Phys. Rev. Lett., **90**, 146801 (2003); X. Cartoixá, D. Z-Y Ting and Y-C Chang, "A resonant spin lifetime transistor", Appl. Phys. Lett., **83**, 1462 (2003).

[8] E. H. Hall, "On a new action of the magnet on electric currents", Am. J. Math., **2**, 287 (1879).

[9] J. Inoue and H. Ohno, "Taking the Hall effect for a spin", Science, **309**, 2004 (2005).

[10] L. Meier, G. Salis, I. Shorubalko, E. Gini, S. Schön and K. Ensslin, "Measurement of Rashba and Dreselhaus spin-orbit magnetic fields", Nature Phys., **3**, 650 (2007).

[11] H. A. Engel, E. I. Rashba and B. I. Halperin, "Out-of-plane spin polarization from in-plane electric and magnetic fields", Phys. Rev. Lett., **98**, 036602 (2007).

[12] J. Sinova, D. Culcer, Q. Niu, N. A. Sinitsyn, T. Jungwirth and A. H. MacDonald, "Universal intrinsic Spin Hall Effect", Phys. Rev. Lett., **92**, 126603 (2004).

[13] S. Murakami, N. Nagaosa and S-C Zhang, "Dissipationless quantum spin current at room temperature", Science, **301**, 1348 (2003).

[14] J. E. Hirsch, "Spin Hall Effect", Phys. Rev. Lett., **83**, 1834 (1999).

[15] Y. K. Kato, R. C. Myers, A. C. Gossard and D. D. Awschalom, "Observation of the Spin Hall Effect in semiconductors", Science, **306**, 1910 (2004).

[16] K. v. Klitzing, G. Dorda and M. Pepper, "New method for high-accuracy determination of the fine structure constant based on quantized Hall resistance", Phys. Rev. Lett., **45**, 494 (1980).

[17] S. Murakami, N. Nagaosa and S-C Zhang, "Spin Hall insulator", Phys. Rev. Lett., **93**, 156804 (2004).

[18] C. L. Kane and E. J. Mele, "Z_2 topological order and the quantum spin Hall effect", Phys. Rev. Lett., **95**, 146802 (2005).

[19] B. A. Bernevig and S-C Zhang, "Quantum spin Hall effect", Phys. Rev. Lett., **96**, 106802 (2006).

[20] B. A. Bernevig, T. L. Hughes and S-C Zhang, "Quantum spin Hall effect and topological phase transition in HgTe quantum wells", Science, **314**, 1757 (2006).

[21] M. König, S. Wiedmann, C. Brüne, A. Roth, H. Buhmann, L. W. Molenkamp, X-L Qi and S-C Zhang, "Quantum spin Hall insulator state in HgTe quantum wells", Science, **318**, 766 (2007).

8

Spin Relaxation

In Chapter 5, we were introduced to the concept of spin relaxation and the T_1 and T_2 times. Spin relaxation is of paramount importance in spintronics. Here, one is often concerned with using the spin polarization of either a single charge carrier (or the net spin polarization of an ensemble of charge carriers) to encode information. An extreme example of this strategy is the Single Spin Logic (SSL) paradigm discussed in Chapter 13. If spin were to host information reliably, it must be protected against random and spontaneous depolarization, caused by "spin relaxation". In this Chapter, we discuss the various mechanisms for spin relaxation in a solid.

When an electron is introduced in a solid, interaction with the environment can affect its spin orientation. The environment may give rise to an *effective magnetic field* which interacts with the spin and causes it to alter its state, thereby changing the orientation. The effective magnetic field can arise from a multitude of sources, e.g., the spins of other electrons and holes in the solid, nuclear spins, phonons (vibrating atoms) that give rise to a time varying magnetic field in some circumstances, and, most importantly, spin-orbit interactions in the solid that act like an effective magnetic field, as we saw in the previous Chapter.

Any effective magnetic field interacts with the magnetic moment of the electron's spin. The interaction energy is given by Equation (6.6). This interaction can alter the electron's spin state. To understand how this occurs, first consider the time-independent Pauli Equation

$$[H_0+H_{so}][\psi] = [H_0-g\mu_B\vec{B}_{eff}\cdot\vec{s}][\psi] = [H_0+(g/2)\mu_B\vec{B}_{eff}\cdot\vec{\sigma}][\psi] = E[\psi] \quad (8.1)$$

describing this electron. The spin-orbit interaction Hamiltonian is written as a Zeeman interaction term, $(g/2)\mu_B\vec{B}_{eff}\cdot\vec{\sigma}$, due to the effective magnetic field arising from the interaction. Assume, for simplicity, that the magnetic field is directed along the z-direction. In that case, the time independent Pauli Equation reduces to

$$\begin{aligned}[H_0 - (g/2)\mu_B B\sigma_z][\psi] &= \begin{bmatrix} H_O - (g/2)\mu_B B & 0 \\ 0 & H_O + (g/2)\mu_B B \end{bmatrix}[\psi] \\ &= E[\psi], \end{aligned} \quad (8.2)$$

where $[\psi]$ is the 2-component wavefunction.

Diagonalization of the Hamiltonian in the basis of unperturbed states ϕ_0 (unperturbed states are eigenstates of H_0) yields the eigenenergies

$$E_{\pm} =< H_O > \pm(g/2)\mu_B B \tag{8.3}$$

where $H_0 >=< \phi_0|H_0|\phi_0 >$. The corresponding eigenspinors are

$$|1 >= \begin{bmatrix} 0 \\ 1 \end{bmatrix} \tag{8.4}$$

$$|0 >= \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \tag{8.5}$$

which are the -z-polarized and +z-polarized spin states. In other words, the stable (time-invariant) spin polarizations in the presence of the effective magnetic field are those that are parallel or anti-parallel to the magnetic field. No other spin polarization is an eigenstate and stable. Note that a magnetic field makes the spin polarization "bistable" (i.e., only the +z and -z polarized states are stable) and this feature is exploited in the realization of the Single Spin Logic (SSL) family discussed in Chapter 13.

Now consider an electron that finds itself in a different effective magnetic field caused by a magnetic scatterer, or other spins, or a different spin-orbit interaction. If the electron's spin polarization is already parallel or anti-parallel to the new effective magnetic field, then the spin will not change since the spin polarization is already an eigenstate and therefore stable. However, this will be a rare situation. Most of the time, the electron's initial spin will not be parallel or anti-parallel to the new effective magnetic field and therefore the initial spin will *not* be stable (or time invariant) and begin to change with time as soon as the electron encounters the new effective magnetic field. We can actually show, starting from the time dependent Pauli Equation (see Problem 8.1 and also recall Chapter 4 material), that the spin will *precess* about the new effective magnetic field B_{eff} with an angular frequency Ω_{eff} $= g\mu_B B_{eff}/\hbar$, where g is the Landé g-factor of the medium where the electron resides. This is the familiar *Larmor precession* and the time evolution of the spin, undergoing Larmor precession, is described by the equation

$$\frac{d\vec{S}}{dt} = \vec{\Omega}_{eff} \times \vec{S}\,, \tag{8.6}$$

which we derived in Chapter 4.

Now, since the flux density B_{eff}, caused by spin orbit interaction, depends on the electron's velocity $\vec{v}$ or wavevector k, therefore if an electron's velocity or wavevector changes randomly owing to scattering, then two different things happen. First, the axis about which the spin precesses changes direction because the effective magnetic field changes direction, and second, the frequency of Larmor precession changes. These changes occur randomly in time because scattering events are random. As a result, the spin orientation

of an electron changes randomly in time. This leads to spin relaxation. We will talk about this more when we discuss the D'yakonov-Perel' mechanism of spin relaxation.

If spin changes gradually with time, we normally call it "spin relaxation", while if the spin changes suddenly (discretely in time) then we tend to call it a "spin flip". A complete spin-flip is one where an "up-spin" state becomes a "down-spin" state and vice versa. But two such states are mutually *orthogonal* and therefore we normally expect no coupling between them since the matrix element connecting these two states should be zero. So, what can couple two orthogonal states and cause transitions between them, resulting in a spin flip? The "scatterer" that couples them and causes transitions must have an internal magnetic field of some sort. In that case, the "up-spin" state acquires a little admixture of the "down-spin" state, and vice versa, so that the two states are no longer completely orthogonal. This will then result in some coupling between them.

A scatterer with an internal magnetic field can also make an electron originally in an eigenspinor state become unstable and change spin orientation. This happens because when we add the scatterer's magnetic field (vectorially) to the original magnetic field that the electron was residing in, we get a new magnetic field. The electron's spin polarization is no longer parallel or anti-parallel to this new magnetic field. Therefore, the electron is no longer in a spin eigenstate. Thus, the scatterer makes the spin unstable and causes it to change.

In the following sections, we discuss the major spin flip mechanisms in a solid. But first, we will revisit the issue of the spin-orbit field B_{eff} (or B_{so}) being proportional to either velocity v or wavevector k. We showed in the previous chapter that the B_{so} which is proportional to k is spin-dependent. Now, we will show that we can define a *spin-independent* B_{so} for either Rashba or Dresselhaus interaction which will be proportional to the *average* wavevector k_{av} in the two spin-split bands. This average wavevector is the arithmetic mean of the wavevectors in the two spin-split bands and is therefore spin-independent.

8.1 The spin-independent spin-orbit magnetic field

In Chapter 7 (Section 7.6), we applied the Ehrenfest Theorem to deduce the spin-orbit field B_{so} from the spin-orbit splitting energy and came up with Equation (7.114). This approach yields that B_{so} is proportional to k, not v. However, that also immediately makes this B_{so} *spin-dependent* since an electron with energy E has two different wavevectors k_1 and k_2 in the two spin-split bands in the presence of spin-orbit interaction. This is shown in

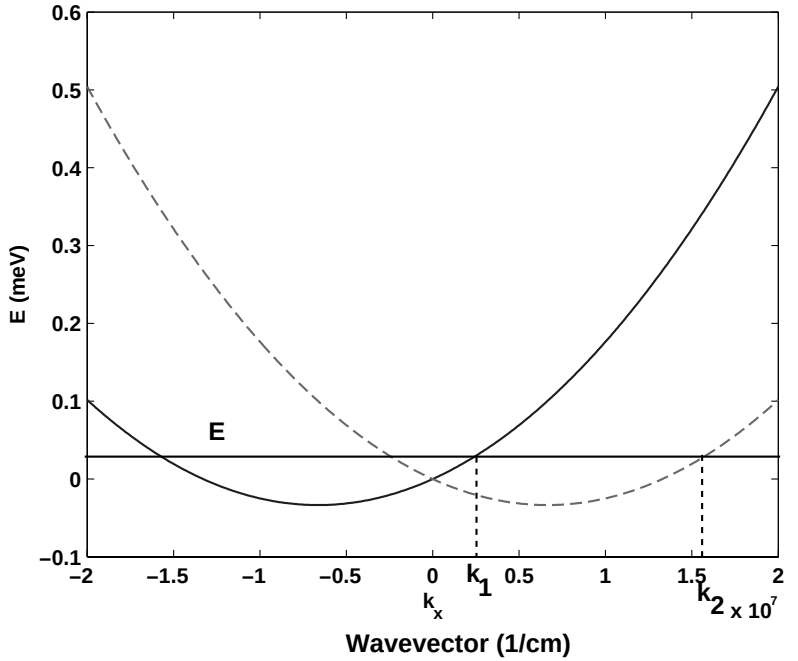


FIGURE 8.1
Dispersion relations of spin-split subbands in a quantum wire with just Rashba interaction. In this plot, we assumed that $\eta = 10^{-11}$ eV-m and $m^* = 0.05$ m_0.

Fig. 8.1.

To define a *spin-independent* or *spin-averaged* B_{so}, we proceed as follows. The dispersion relations in the presence of either Rashba or Dresselhaus interaction in a one- or two-dimensional electron gas is (neglecting band structure non-parabolicity and any external magnetic field):

$$E^{\pm}(k) = \frac{\hbar^2 k^2}{2m^*} \pm \eta k \quad (RASHBA)$$
$$E^{\pm}(k) = \frac{\hbar^2 k^2}{2m^*} \pm \nu k \quad (DRESSELHAUS). \tag{8.7}$$

These are the dispersion relations shown in Fig. 8.1.

The velocities of an electron with energy E in the two spin-split bands are

$$\begin{aligned} v_1(E) &= \frac{1}{\hbar}\frac{\partial E^+}{\partial k}\bigg|_{E=E^+} = \frac{\hbar k_1(E)}{m^*} + \eta \quad (RASHBA) \\ &= \frac{1}{\hbar}\frac{\partial E^+}{\partial k}\bigg|_{E=E^+} = \frac{\hbar k_1(E)}{m^*} + \nu \quad (DRESSELHAUS), \\ v_2(E) &= \frac{1}{\hbar}\frac{\partial E^-}{\partial k}\bigg|_{E=E^-} = \frac{\hbar k_2(E)}{m^*} - \eta \quad (RASHBA) \\ &= \frac{1}{\hbar}\frac{\partial E^-}{\partial k}\bigg|_{E=E^+} = \frac{\hbar k_2(E)}{m^*} - \nu \quad (DRESSELHAUS). \end{aligned} \tag{8.8}$$

From Equation (8.7), we get

$$\begin{aligned} k_2(E) - k_1(E) &= \frac{2m*\eta}{\hbar} \quad (RASHBA) \\ &= \frac{2m*\nu}{\hbar} \quad (DRESSELHAUS). \end{aligned} \tag{8.9}$$

Using the last equation in Equation (8.8), we get

$$v_1(E) = v_2(E) = v(E), \tag{8.10}$$

which shows that the velocity of an electron with energy E is the same in both spin-split bands, so that the velocity is *spin-independent*. Moreover, from Equation (8.8), we obtain

$$v(E) = \frac{v_1(E) + v_2(E)}{2} = \frac{\hbar}{m^*}\frac{k_1(E) + k_2(E)}{2} = \frac{\hbar k_{av}}{m^*}. \tag{8.11}$$

Since the *spin-averaged* B_{so} should be proportional to spin-averaged k, i.e., k_{av}, we can posit that the spin-averaged B_{so} should be proportional to v since v and k_{av} have turned out to be proportional to each other. Since v is spin-independent, the spin averaged B_{so} will also be spin-independent, as it should be.

To find an expression for the spin-independent B_{Rashba}, we should replace k in Equation (7.114) with $m^*v/\hbar$. This will yield

$$|B_{Rashba}| = \frac{2\eta m*}{g\mu_B\hbar}v = \frac{2^{3/2}\sqrt{m^*}\eta}{g\mu_B\hbar}\sqrt{E} \tag{8.12}$$

Similarly,

$$|B_{Dresselhaus}| = \frac{2\nu m^*}{g\mu_B\hbar}v = \frac{2^{3/2}\sqrt{m^*}\nu}{g\mu_B\hbar}\sqrt{E}. \tag{8.13}$$

Note that the above equations clearly show that the spin-independent magnetic fields depend only on the electron energy regardless of the spin orientation.

In the rest of this book, we will use the spin-independent magnetic fields in Equations (8.12) and (8.13) whenever we talk about spin precession about an effective magnetic field caused by spin-orbit interaction. In that case, we will not have to worry about the effective magnetic field itself changing with the precessing spin. The notion of the spin-independent effective magnetic field associated with spin-orbit interaction allows us to present a completely classical description of the Spin Field Effect Transistors operation in Chapter 12.

8.2 Spin relaxation mechanisms

The four major spin relaxation mechanisms of electrons in the conduction band of a semiconductor are (i) the D'yakonov-Perel' mechanism, (ii) Elliott-Yafet mechanism, (iii) the Bir-Aronov-Pikus mechanism, and (iv) hyperfine interactions with nuclear spins. Physical pictures to aid in the understanding of the first three mechanisms were presented in a paper by Fabian and Das Sarma [1].

8.2.1 Elliott-Yafet mechanism

The Elliott-Yafet mechanism is a result of the fact that in a crystal, the Bloch states (which are solutions of the Schrödinger equation in the periodic lattice potential) may not be spin eigenstates. As a result, an electron's spin in the crystal does not really have one of two fixed polarizations - "up" or "down", contrary to what we normally assume. The actual polarization depends slightly on what the electron's wavevector is. This is schematically depicted in Fig. 8.2 where we plot the energy-wavevector relation in an arbitrary band and show that the spin orientation of an electron (depicted by the arrow) is slightly different in different wavevector states. Each wavevector state still has two possible spin orientations that are mutually anti-parallel, but spin orientations associated with different wavevector states can have arbitrary angles between them. This is a situation that we encountered in the previous Chapter, where it was caused by an external magnetic field. Here, it is caused by the fact that the Bloch eigenstates are not spin eigenstates in a crystal that lacks inversion symmetry.

If a collision event with a non-magnetic scatterer (such as an impurity, device boundary, or phonon) changes the momentum or wavevector of an electron, then it will change the spin as well since the spin orientations associated with the initial and final wavevector states are never mutually parallel. This process is also shown in Fig. 8.2. By how much the spin orientation changes naturally depends on how much the wavevector changes. Collisions

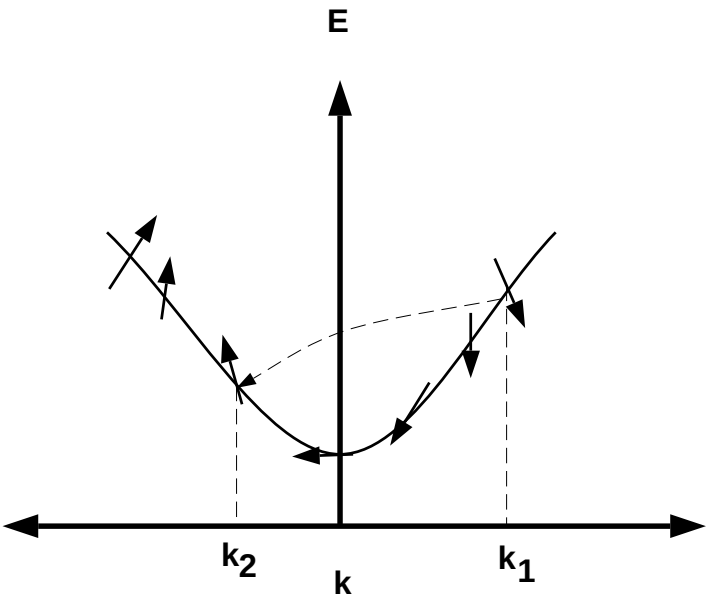


FIGURE 8.2
Energy dispersion relation showing the spin polarizations at different wavevector states. Each wavevector state has two possible mutually anti-parallel spin polarizations, one of which is shown by the arrowheads in the above figure. A momentum changing collision event, that changes the wavevector state from k_1 to k_2, also changes the spin orientation since the eigenspin polarizations at k_1 and k_2 are generally not parallel.

with charged impurities that do not usually change the wavevector by a lot* will not be very effective in changing spin orientation this way, but collisions with certain types of acoustic phonons that prefer scattering through large angles (and therefore changes the wavevector by a lot) will be very effective in relaxing spin.

Elliott [2] had considered an even more drastic situation. The spin state of a nearly up-spin electron at a wavevector state k_1 and that of a nearly down-spin electron at a *different* wavevector state k_2 are *not* strictly orthogonal, since they are not strictly anti-parallel. Thus, any collision with a non-magnetic scatterer that changes the wavevector of an electron from an initial state k_1 to a state k_2 can also couple the nearly down-spin state at k_1 to the nearly up-spin state at k_2. This coupling can flip an electron's spin from nearly up to nearly down. Note that this kind of spin relaxation is always accompanied by some degree of momentum relaxation since the wavevector must change in order to change the spin.

Reference [3] has calculated an approximate expression for the Elliott-Yafet spin relaxation rate in bulk semiconductors assuming that the carriers are in a non-degenerate population, in quasi equilibrium, and are described by a thermal distribution with temperature T. This expression is

$$\frac{1}{\tau_{EY}} = A\left(\frac{k_B T}{E_g}\right)^2 \mathcal{R}^2 \left(\frac{1-\mathcal{R}/2}{1-\mathcal{R}/3}\right)^2 \frac{1}{\tau_m}, \tag{8.14}$$

where E_g is the bandgap, $\mathcal{R} = \Delta/(E_g + \Delta)$, Δ is the spin-orbit splitting of the valence band, τ_m is the momentum relaxation time, k_B is the Boltzmann constant, and A is a constant varying from 2 to 6 depending on the dominant scattering mechanism for momentum relaxation.

This expression shows that the Elliott-Yafet spin relaxation rate $1/\tau_{EY}$ is directly proportional to the momentum relaxation rate $1/\tau_m$. This is to be expected since Elliot-Yafet relaxation is *caused* by momentum relaxation. However, what this expression also shows is that the spin and momentum relaxation rates have *different* temperature dependences. For example, if the momentum relaxation rate is independent of temperature (as expected for mostly elastic scattering events), then the spin relaxation rate should roughly increase quadratically with temperature since the temperature dependence of bandgap and spin-orbit interaction strengths are negligible. However, in a metal or strongly degenerate electron gas, the Elliott-Yafet spin relaxation rate and the momentum relaxation rate tend to have the same temperature dependence. Elliott derived a relation between the two rates as [4]

$$\frac{\tau_m}{\tau_{EY}} \propto \frac{\Delta}{E_g}, \tag{8.15}$$

*The matrix element for electron-impurity interaction typically has an inverse dependence on the net momentum change, meaning that small angle scatterings are preferred over large angle scatterings.

which Yafet has shown to be temperature-independent [5]. According to the above equation, the two relaxation rates must have the same temperature dependence if we ignore the weak temperature dependences of Δ and E_g. There is also some experimental evidence that indeed the two rates have the same temperature dependence in metals [6].

The Elliott-Yafet mechanism is often the primary spin relaxation mechanism in low mobility materials, such as organics, where momentum relaxing scattering events are frequent. In high mobility semiconductors, the primary spin relaxation mechanism is usually the D'yakonov-Perel' mechanism which we discuss next.

8.2.2 D'yakonov Perel' mechanism

If a solid lacks inversion symmetry because of its crystalline structure (such as in compound semiconductors) or because of an externally applied electric field (or a built-in electric field), then an electron in the solid will experience strong spin-orbit interaction. As shown in Chapter 6, the crystallographic inversion asymmetry (also known as bulk inversion asymmetry) gives rise to the Dresselhaus spin-orbit interaction, while the structural inversion asymmetry caused by an external (or built-in) electric field gives rise to the Rashba spin-orbit interaction. These interactions lift the degeneracy between up-spin and down-spin states at any non-zero wavevector (recall Chapters 6-7), so that these two states have different energies in the same wavevector state. In this respect, Dresselhaus and Rashba spin-orbit interactions truly behave like effective magnetic fields, since a magnetic field also lifts the degeneracy between up-spin and down-spin states at any given wavevector because of the Zeeman interaction. The effective magnetic field $\vec{B}(\vec{v})$, arising from either Rashba or Dresselhaus interaction, will be dependent on the electron's velocity $\vec{v}$ if it is spin-independent[†].

Earlier, we mentioned that the effective magnetic field causes an electron's spin to undergo Larmor precession. The precession axis is collinear with the magnetic field and the precession frequency is $\vec{\Omega}(\vec{v}) = e\vec{B}(\vec{v})/m^*$. Now consider an ensemble of electrons drifting and diffusing in a solid. If every electron's velocity $\vec{v}$ is the same and does not change with time, then the field $\vec{B}(\vec{v})$ is the same for all electrons and every electron precesses about this constant field with a fixed frequency $|\vec{\Omega}(\vec{v})|$. This does *not* cause any spin relaxation at all. If all electrons started with the same spin polarization, then after any arbitrary time, they all have precessed by exactly the same angle, and therefore they all have again the same spin polarization. The direction of this spin polarization may be different from the initial one, but that does not matter. The *magnitude* of the ensemble averaged spin does not change with time. Therefore, there is no spin relaxation.

[†]See the discussion in the Section 8.1.

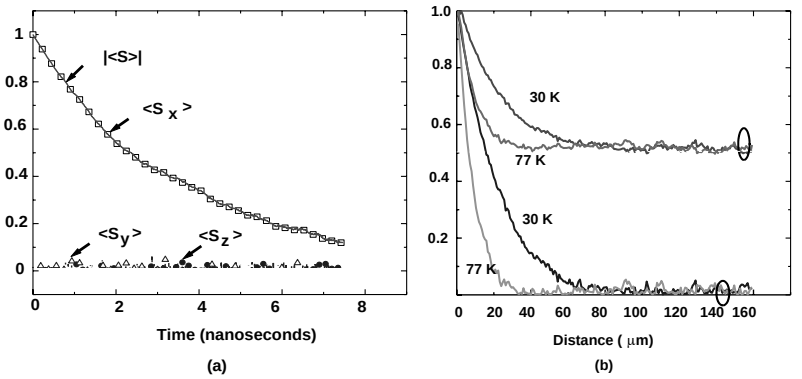


FIGURE 8.3
Decay of ensemble averaged spin components $< S_x >$, $< S_y >$ and $< S_z >$, as well as $| < S > | = \sqrt{< S_x >^2 + < S_y >^2 + < S_z >^2}$, in a GaAs quantum wire of width 30 nm and thickness 4 nm due to D'yakonov-Perel' relaxation. The strength of the Rashba interaction a_{46} is assumed to be 10^{-38} C-m^2 and the strength of the Dresselhaus interaction is assumed to be $a_{42} = 2 \times 10^{-29}$ ev-m^3. There is an electric field of 2 kV/cm along the wire's length, driving transport. Relaxation-in-time plot (a) is reprinted with permission from S. Pramanik, et al., Phys. Rev. B, **68**, 075313 (2003). Copyright (2003) American Physical Society. Relaxation-in-space (distance along the length of the quantum wire) plot (b) is reprinted with permission from S. Pramanik, et al., Appl. Phys. Lett., **84**, 266 (2004). Copyright (2004) American Institute of Physics.

However, if $\vec{v}$ changes *randomly* with time because of scattering, then different electrons would have precessed by different angles after a certain time because they have different scattering histories. Thus, if all electrons were injected with the same spin polarization, their spin polarizations gradually go out of phase with each other. After a sufficiently long time, the magnitude of the ensemble averaged spin will decay to zero. This is the basis of D'yakonov-Perel' (D-P) spin relaxation.

It is obvious that $\vec{B}(\vec{v})$ – the velocity dependent effective magnetic field arising from Dresselhaus and Rashba spin-orbit interactions – is the cause of D'yakonov-Perel' relaxation. Therefore, anything that reduces $\vec{B}(\vec{v})$ will suppress D-P relaxation. Slower moving electrons experience a smaller $\vec{B}(\vec{v})$ since the magnitude of $\vec{B}(\vec{v})$ is proportional to the magnitude of $\vec{v}$. Thus, they are less susceptible to D-P relaxation. One would therefore expect that everything else being equal, high carrier mobility, or high saturation velocity of carriers, *reduces* spin lifetimes. That may indeed be true in most cases.

Therefore, frequent momentum relaxing collisions may actually be beneficial to spin longevity. These collisions will tend to slow down electrons, reduce their average momentum, and therefore suppress D-P relaxation by reducing $\vec{B}(\vec{v})$ as originally noted by D'yakonov and Perel'. This phenomenon has come to be known as "motional narrowing". On the other hand, very strong momentum relaxing mechanisms also tend to put the spin polarizations of different electrons quickly out of phase with one another by causing a larger spread in the velocity of the carriers. Thus, it is not clear *a priori* whether momentum relaxing collisions are ultimately beneficial or detrimental to spin lifetimes. The answer may be different in different systems[‡].

The existence of the motional narrowing effect has been demonstrated indirectly in a remarkable experiment by Sandhu et al. [8]. They studied spin relaxation in a two-dimensional electron gas (a quantum well) which had a metal gate placed on top of the well. By applying an electrostatic potential to the gate, they could change the carrier concentration in the well and study the spin relaxation time as a function of carrier concentration. At high carrier concentration, the impurities in the well are screened by the electrons. As a result, the momentum relaxation rate due to impurity scattering is suppressed. Therefore, the D'yakonov-Perel' mode dominates. At lower carrier concentration, the screening is weakened and the momentum relaxation rate goes up. This reduces the D'yakonov-Perel' rate and the dominant spin relaxation mode becomes the Bir-Aronov-Pikus mode discussed in the next section. Sandhu's experiment was presumably carried out at small biases where hot carrier effects are unimportant and the motional narrowing effect is clearly manifested.

Reference [3] has derived an approximate expression for the spin relaxation rate due to the D'yakonov-Perel' mechanism in bulk *non-degenerate* semiconductors where the carriers are in quasi equilibrium described by Boltzmann statistics. Furthermore, the derivation assumes that the primary source of $\vec{B}(\vec{v})$ is the Dresselhaus interaction. Pikus' expression is

$$\frac{1}{\tau_{DP}} = Q\alpha^2 \frac{(kT)^3}{\hbar^2 E_g} \tau_m, \tag{8.16}$$

[‡]Many researchers believe that more frequent momentum relaxing scattering leads to longer D-P spin relaxation times. They think that the mean of the velocity distribution, which determines the average $\vec{B}(\vec{v})$, is determined by the frequency of momentum relaxation events, whereas the standard deviation of the velocity distribution is determined solely by temperature and has nothing to do with the momentum relaxation rate. This may be correct if the electron system is in quasi equilibrium and described by Boltzmann statistics. However, if the system is driven far out of equilibrium and experiences hot carrier effects, then this argument is invalid. Our simulations [7] have shown that there are two regimes of transport: the "motional narrowing regime" where frequent momentum relaxation events suppress D-P relaxation and the "inhomogeneous broadening regime" where frequent momentum relaxing scattering enhances D-P relaxation. The inhomogeneous broadening regime seems to appear only in high field transport where hot carrier effects are important.

where Q is a dimensionless quantity ranging from 0.8 - 2.7 depending on the dominant momentum relaxation process, E_g is the bandgap, τ_m is the momentum relaxation time, and α is a measure of the Dresselhaus interaction strength given by

$$\alpha \approx \frac{4\eta}{\sqrt{3-\mathcal{R}}}\frac{m^*}{m_0}, \tag{8.17}$$

where m_0 is the free electron mass, m^* is the electron's effective mass and $\mathcal{R}$ was defined previously in connection with the Elliott-Yafet mechanism. Note that this expression shows that the spin relaxation rate is inversely proportional to the momentum relaxation rate which is true only in the motional narrowing regime. Furthermore, it shows that the spin relaxation rate will have a strong temperature dependence even if the momentum relaxation rate is temperature independent.

The temperature dependence of the D-P relaxation is important. In a strongly degenerate semiconductor, or metal, with large carrier concentration, the Fermi energy E_F is much larger than the thermal energy kT. Therefore, the velocity of electrons carrying current is largely determined by E_F and not kT. In that case, the D-P relaxation rate is more or less independent of temperature. By the same consideration, if the drift velocity of electrons is much larger than the thermal velocity, then the D-P relaxation will be again more or less temperature independent. Otherwise, the D-P relaxation will depend on the temperature. The exact dependence will depend on the temperature dependence of the momentum relaxing collisions which affect the carrier momentum. Generally, the D-P relaxation rate tends to increase slowly with increasing temperature [9].

8.2.2.1 D-P relaxation in a quantum wire

A very interesting question is whether there will be D-P relaxation in quantum wires? In a quantum wire (quasi one-dimensional system), $\vec{v}$ is always along the axis of the wire and therefore cannot change *direction* because of scattering or acceleration due to electric field (although it can change magnitude). Now, the spin-independent field $\vec{B}(\vec{v})$ depends on $\vec{v}$, but is *not* collinear with $\vec{v}$. In a quantum wire, the contribution to $\vec{B}(\vec{v})$ from the Dresselhaus interaction is indeed collinear with $\vec{v}$, but the contribution from the Rashba interaction is perpendicular to $\vec{v}$. Consider a quantum wire with rectangular cross-section shown in Fig. 8.4. The wire axis is along the x-direction and the symmetry breaking electric field which induces the Rashba spin-orbit interaction is along the y-direction. Current flows along the x-direction which is also assumed to be the crystallographic [100] direction. In this system, the Dresselhaus and Rashba contributions to $\vec{B}(\vec{v})$ will be given by (see Problem 8.2)

$$\vec{B}_{Dresselhaus}(\vec{v}) = \frac{2m^* a_{42}}{g\mu_B\hbar}\left[\left(\frac{m\pi}{W_z}\right)^2 - \left(\frac{n\pi}{W_y}\right)^2\right] v_x\hat{x}$$

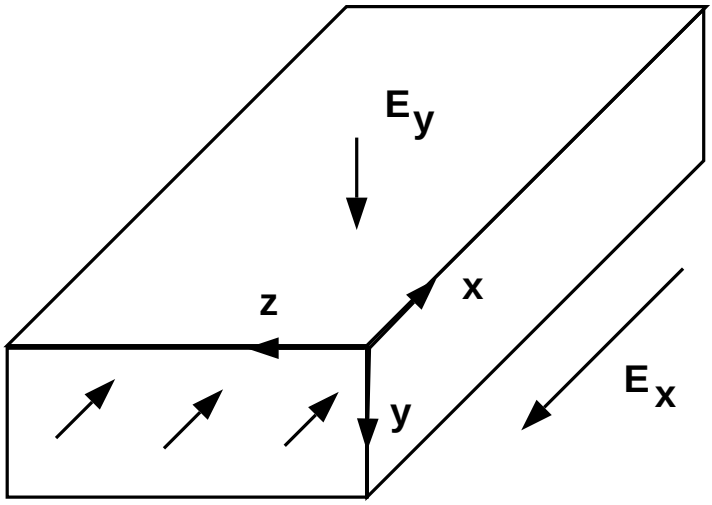


FIGURE 8.4
A quantum wire with rectangular cross-section. Electrons are injected at $x = 0$, with 100% spin injection efficiency.

$$\vec{B}_{Rashba}(\vec{v}) = \frac{2m^* a_{46}}{g\mu_B \hbar} E_y v_x \hat{z}, \tag{8.18}$$

where W_z, W_y are the transverse dimensions of the wire, E_y is the symmetry-breaking electric field that induces the Rashba effect, a_{42}, a_{46} are material constants, v_x is the electron's velocity along the wire axis, and m, n are the transverse subband indices for the subband in which the electron resides. Note that the effective magnetic field due to the Dresselhaus interaction is along the wire axis, but that due to the Rashba interaction is mutually perpendicular to the wire axis and the symmetry breaking electric field. The total magnetic field $\vec{B}(\vec{v})$ $(= \vec{B}_{Rashba}(\vec{v}) + \vec{B}_{Dresselhaus}(\vec{v}))$ is therefore in the x-z plane and subtends an angle α with the wire axis given by

$$\alpha = arctan\left(\frac{B_{Rashba(\vec{v})}}{B_{Dresselhaus}(\vec{v})}\right) = arctan\left(\frac{a_{46}E_y}{a_{42}\left[\left(\frac{m\pi}{W_z}\right)^2 + \left(\frac{n\pi}{W_y}\right)^2\right]}\right). \tag{8.19}$$

Note that the angle α is independent of the velocity. Now, if the electron

is always in a fixed subband defined by a fixed set of indices m, n, then the angle α (and therefore the axis of spin precession) will be *fixed* and invariant in time. In other words, the *direction* of the effective magnetic field $\vec{B}(\vec{v})$ does not change with time, although the *magnitude* does. Therefore, every electron precesses about the *same* axis, albeit with different frequencies. So, will there be any D-P spin relaxation? The correct answer is extremely intriguing. There will be no D-P relaxation in *space* [10], but there will be D-P relaxation in *time*! We show this in Problem 8.3. Thus, there is a fundamental *difference* between spin relaxation in space versus spin relaxation in time. There is no analog of this in momentum or energy relaxation. Spin is indeed unique in this sense.

If there is inter-subband scattering and the electrons change subbands, then the angle α will change every time a subband transition occurs. In that case, both the magnitude *and* the direction of $\vec{B}(\vec{v})$ change randomly in time. In this case, there will be D-P relaxation in both time and space. Note also, that neither Rashba nor Dresselhaus interaction *alone* can cause spatial D-P relaxation in a quantum wire even if inter-subband scattering occurs. Both interactions have to be present. If either one is absent, then α is either 0 or $\pi/2$, i.e., α is fixed and D-P relaxation in space does not occur. Thus, the two necessary ingredients for *spatial* D-P relaxation in a quantum wire are (i) simultaneous presence of both Dresselhaus and Rashba interactions, and (ii) multichanneled transport. This is shown again in Problem 8.3.

Even if transport is not strictly single channeled, the D-P relaxation is progressively subdued in a quantum wire as its transverse dimensions shrink. This has been theorized in a number of publications [11, 12]. There is also some experimental evidence to support this postulate [13]. Thus, quantum wires have an advantage over quantum wells when it comes to spin relaxation. Spin relaxation via the D-P mode can be dramatically suppressed in quantum wires. As a result, many of the spin transistors described in Chapter 12, which are adversely affected by D-P spin relaxation, are best implemented with quantum wire channels as opposed to quantum well channels.

Another important issue that has lately received some attention is the anisotropy of spin relaxation. Because the effective magnetic fields due to the Dresselhaus and Rashba interactions point in different directions, as noted above, the interplay between these two interactions will make the spin relaxation rate depend on the initial polarization of the spin, i.e., what angle the initial spin polarization subtends with the net effective magnetic field caused by Rashba and Dresselhaus interactions. This has been noted in both quantum wires [14] and quantum wells. The Dresselhaus interaction also depends strongly on the crystallographic orientation (Miller indices) associated with the direction of spin transport. This can give rise to anisotropy of spin relaxation rates of a different kind, meaning that spin relaxation rate will depend strongly on the direction of spin transport within a crystal. This has been studied theoretically in the context of quantum wells [15, 16]. The anisotropy

of spin relaxation rates can be quite strong. The relaxation rate for spins originally polarized along one particular crystallographic direction may differ from that originally polarized along another crystallographic direction by a few orders of magnitude! A dramatic experimental demonstration of the influence of crystallographic orientation was shown by Ohno, et al. [17]. They showed that the measured spin relaxation time in (110) GaAs quantum wells is an order of magnitude larger than in (100) GaAs quantum wells. Therefore, when it comes to implementing spintronic devices where spin relaxation is a major impediment to good performance, it pays to choose the crystallographic orientation of the device structure judiciously in order to improve device characteristics.

8.2.3 Bir-Aronov-Pikus mechanism

The Bir-Aronov-Pikus mechanism is a source of spin relaxation of electrons in semiconductors where there is a significant concentration of both electrons and holes. In that case, an electron will be usually in close proximity to a hole, so that their wavefunctions will overlap, which will cause an exchange interaction between them. Exchange interaction is the subject of the next chapter. If the hole's spin flips, then exchange interaction will cause the electron's spin to flip as well, leading to spin relaxation.

Reference [3] has calculated approximate expressions for the spin relaxation rate of electrons due to the Bir-Aronov-Pikus mechanism in bulk semiconductors. For a non-degenerate semiconductor, Pikus' expression is

$$\frac{1}{\tau_{BAP}} = \frac{2a_B^3}{\tau_0 v_B}\left(\frac{2E}{m^*}\right)^{1/2}\left[n_f|\psi(0)|^4 + \frac{5}{3}n_b\right], \tag{8.20}$$

where n_a (n_b) is the concentration of free (bound) holes, E is the electron energy in the conduction band, m^* is the electron effective mass, and

$$\begin{aligned}
\tau_0 &= \frac{64E_B\hbar}{3\pi\Delta_x^2},\\
a_B &= \left(\frac{m_0}{\mu}\right)\epsilon_0 a_0,\\
v_B &= \frac{\hbar}{\mu a_B},\\
E_B &= \frac{\hbar^2}{2\mu a_B^2},\\
|\psi(0)|^2 &= \frac{2\pi}{\kappa}\left(1 - e^{-2\pi/\kappa}\right)^{-1},\\
\kappa &= \sqrt{E/E_B},
\end{aligned} \tag{8.21}$$

where a_o is the Bohr radius of Hydrogen atom (0.529 Å), ϵ_0 is the dielectric constant of vacuum, μ is the electron-hole reduced mass, and Δ_x is the

exchange splitting of the exciton ground state. For a semiconductor with degenerate concentration of holes, the result is

$$\begin{aligned}\frac{1}{\tau_{BAP}} &= \frac{2a_B^3}{\tau_0 v_B}\left(\frac{E}{E_F}\right) n_f |\psi(0)|^4 (2E/m^*)^{1/2}; \quad if\ E_F < E(m_v/m^*) \\ &= \frac{2a_B^3}{\tau_0 v_B}\left(\frac{E}{E_F}\right) n_f |\psi(0)|^4 (2E_F/m^*)^{1/2}; \quad if\ E_F > E(m_v/m^*)\end{aligned} \tag{8.22}$$

where E_F is the Fermi energy and m_v is the relevant hole effective mass (light- or heavy-hole, whichever is participating in the Bir-Aronov-Pikus mechanism).

8.2.4 Hyperfine interactions with nuclear spins

In a solid, the nuclear spins generate a magnetic field which interacts with the electron spins via hyperfine contact interaction and can cause spin relaxation. The Hamiltonian describing this interaction is given by

$$H_{nuclear} = \vec{S} \cdot \sum_i A_i \vec{I}_i, \tag{8.23}$$

where $\vec{I}_i$ is the spin of the i-th nucleus, A is a constant and A_i is the corresponding coupling coefficient. The quantity A_i is given by

$$A_i = AV|\psi(\vec{r}_i|^2, \tag{8.24}$$

where $\psi(\vec{r}_i)$ is the electron envelope wavefunction at the nuclear site $\vec{r}_i$ and V is the volume of the crystal cell.

In order to study how the electron spin decays in time, one needs to evaluate the correlator [18]

$$C_n(t) = < n|[S_z(t) - S_z(0)]S_z(0)|n >, \tag{8.25}$$

where $|n >$ is the coupled electron and nuclear state, $S_z(0)$ is the initial spin eigenstate at time $t = 0$ and $S_z(t)$ is the eigenstate at some later time t given by

$$e^{iH_{nuclear}t} S_z(0) e^{-iH_{nuclear}t}. \tag{8.26}$$

From the behavior of $C_n(t)$, one can deduce the temporal behavior of $S_z(t)$.

Khaetski, et al. [18] have studied this problem for both unpolarized and fully polarized nuclei in a quantum dot. In the former case, the ensemble averaged spin eigenstate $< S_z(t) >$ oscillates between $S_z(0)$ and -$S_z(0)$ with a period $\sqrt{N}/A$ where N is the number of nuclear spins interacting with the electron. This behavior persists up to a time $\sim N/A$. Then $< S_z(t) >$ decays non-exponentially with time. In the case of a fully polarized nucleus, $< S_z(t) >$ initially oscillates with a period $1/A$ up to a time $\sim N/A$ and then

it decays non-exponentially with time. The decay is much faster for the fully polarized case, but the steady state value $< S_z(\infty) >$ is much closer to $S_z(0)$ in the fully polarized case than in the unpolarized case. It is interesting to note that in neither case $< S_z(\infty) > = 0$, meaning that hyperfine interaction does not cause complete loss of electron spin polarization.

8.3 Spin relaxation in a quantum dot

We will pay special attention to the spin relaxation of an electron in a quantum dot since this system forms the basis of two important technologies: single spin logic (SSL) and spin-based quantum logic gates. These are discussed in Chapters 13 and 14. In a quantum dot hosting a single electron, there are *five* major sources of spin relaxation. These are (i) spin-orbit interaction arising from any local electric field that causes Rashba type interaction, (ii) spin-orbit interaction associated with crystallographic inversion asymmetry in the quantum dot material, which is reminiscent of the Dresselhaus interaction, (iii) spin orbit interaction caused by strain field in the solid generated by phonons, (iv) spin phonon coupling in the presence of an external magnetic field which mixes the valence band and conduction band states, (v) hyperfine interactions with the nuclear spin. All of these effects give rise to essentially effective magnetic fields that can interact with the spin of an electron in the quantum dot and change its state.

The Hamiltonians describing these perturbations are [19]:

$$
\begin{aligned}
H_1 &= \alpha\vec{\mathcal{E}} \cdot [\vec{\sigma} \times \left(\vec{p} + e\vec{A}\right)] \\
H_2 &= \beta\vec{\sigma} \cdot \vec{\kappa} \\
H_3 &= \gamma\vec{\sigma} \cdot \varsigma \\
H_4 &= \hat{g}\mu_B(u_{xz}\sigma_x + u_{yz}\sigma_y)B_z \qquad (8.27) \\
H_5 &= \mathcal{A}\sum_{i,k} \mathbf{S}_i \cdot \mathbf{I}_k \delta\left(\mathbf{r}_i - \mathbf{R}_k\right), \qquad (8.28)
\end{aligned}
$$

where α and β are material constants that depend on the bandgap, spin-orbit splitting energy, and the electron's effective mass; $\vec{\mathcal{E}}$ is a local electric field caused by either a charged impurity or confinement potential; $\vec{\sigma}$ is the Pauli spin matrix; $\vec{p}$ is the momentum operator; $\vec{A}$ is the vector potential associated with any external magnetic field; $\vec{\kappa}$ is a vector whose x-component is given by $\kappa_x = p_x(p_y^2 - p_z^2)$ and the other components are obtained by cyclic permutation of the indices, with x, y, z being the principal crystallographic axes; γ is a material constant that depends on the spin-orbit splitting energy and bandgap of the material; λ is a vector whose x-component is given by ς_x

$= (1/2)[u_{xy}, p_y]_+ - (1/2)[u_{xz}, p_z]_+$, where $[,]_+$ denotes the anti-commutator (the other components are obtained by cyclic permutations of the indices); u_{ij} is the lattice strain tensor; $\hat{g}$ is another material constant (not the electron Landé g-factor) that depends on effective mass, spin-orbit splitting energy in the valence band, bandgap and one of three deformation potential constants describing the strain effect on the hole band splitting; μ_B is the Bohr magneton; B_z is a z-directed magnetic flux density associated with any external magnetic field present (its direction is taken to be the z-direction); $\mathcal{A}$ is the hyperfine constant; while $\vec{S}_i(\vec{I}_k)$ and $\vec{r}_i(\vec{R}_k)$ denote the spin and position of the i-th electron (k-th nuclei).

The presence of the Pauli spin matrix $\vec{\sigma}$ in H_n (n = 1,2,3,4) indicates that unless the electron is initially in an eigenspinor state of this Hamiltonian, it will be coupled to its orthogonal eigenspinor and therefore the Hamiltonian $\sum_{n=1}^{4} H_n$ will cause a spin flip event. Within the Born approximation, this spin flip rate $\Gamma_{\uparrow\downarrow}$ can be found from the Fermi's Golden Rule

$$\Gamma_{\uparrow\downarrow} = \frac{2\pi}{\hbar} | < \psi_\downarrow | \sum_{n=1}^{5} H_n | \psi_\uparrow > |^2 \delta(E_\downarrow - E_\uparrow \pm \hbar\omega), \tag{8.29}$$

where $\psi_\downarrow$ and $\psi_\uparrow$ are the wavefunctions (inclusive of the spatial and spinor parts) of the final and initial states of the electron, $E_\downarrow$ and $E_\uparrow$ are the energies of the final and initial states, and $\hbar\omega$ is the energy of phonon emitted or absorbed in the transition process. In the equation above, the plus sign within the delta-function corresponds to absorption and the minus sign to emission of a phonon.

The spin flip rates have been calculated in different types of quantum dots at various temperatures (see, for example, [19]). These rates are very small, typically less than 10^5/sec in GaAs quantum dots under reasonable conditions. In fact, at very low temperatures, the spin-flip rate in GaAs quantum dots has been measured to be smaller than 1/sec [20]. Therefore, quantum dots are ideal hosts for spin since spin is a relatively long lived entity (lifetime > 10 μs) in these systems.

Many research groups have engaged in experimentally measuring spin flip times in quantum dots. One of the earlier experiments measured the spin flip rates in GaAs quantum dots each populated with a single electron [21]. A magnetic field was applied to split the lowest subband state into two Zeeman sublevels. A short voltage pulse that affects the confining potential of the quantum dot was used to tune the sublevel energies with respect to the chemical potentials in the leads and thus selectively populate the higher sublevel. Relaxation to the lower sublevel was then measured. At a temperature of 20 mK and under an applied magnetic field of 7.5 Tesla, the spin relaxation time in the GaAs quantum dots was found to be larger than 50 μsec. In InP quantum dots, the spin relaxation time exceeds 100 μsec at a temperature of 2 K [22]. More recent measurements have shown relaxation times can reach several milliseconds in GaAs quantum dots [23], 20 milliseconds in InGaAs

quantum dots at a temperature of 1 K [24], and even 170 milliseconds in a GaAs quantum dot [25] at a magnetic field of 1.7 Tesla (later extended to 1 second [20]). But all these pale in comparison with organic nanostructures. The spin relaxation time in 50-nm diameter nanowires of the π-conjugated organic tris(8-hydroxyquinolinolato-aluminum), popularly known as Alq_3[§], was found to be at least several msec, and perhaps even 1 second, at the relatively high temperature of 100 K [26]. This is caused by the extremely weak spin-orbit interactions in organics [27]. InAs quantum dots have much stronger spin-orbit coupling and therefore reported relaxation time in these dots is considerably smaller [28].

The spin relaxation time in various materials is summarized in the Table below [27].

TABLE 8.1
Spin relaxation time in various materials

Material	Spin relaxation time (sec)
Copper (bulk)	5×10^{-13}
Cobalt (bulk)	4×10^{-14}
GaAs (bulk)	10^{-9} - 10^{-7}
GaAs (quantum dot)	10^{-1} - 1
Alq_3 (bulk or nanowire)	10^{-2} - 1

8.3.1 Longitudinal and transverse spin relaxation times in a quantum dot

In Chapter 5, we introduced the notion of longitudinal and transverse spin relaxation times T_1 and T_2. Since spins in quantum dots are often used as vehicles to host qubits, what is often important is not the longitudinal spin relaxation time T_1, but rather the spin coherence time, or the transverse relaxation time T_2. The T_1 time essentially refers to the mean time that elapses before an excited spin flips to the ground state. This time is of great importance in classical spin-based logic gates that we will discuss in Chapter 13, but less relevant to quantum computing where what matters is how rapidly a single spin loses its quantum mechanical phase coherence. That time is determined by the transverse relaxation time, or T_2 time. The definition of the T_2 time invariably involves the concept of quantum superposition. As we saw in Chapter 3, a qubit is a coherent superposition of up-spin and down-spin

[§]This organic is one of the most popular materials used as the electron transport layer in organic light emitting diodes.

states:

$$qubit = \alpha|\uparrow> + \beta|\downarrow> . \tag{8.30}$$

In order to preserve the coherence between the up-spin and down-spin states, one needs to maintain the precise phase relationship between the complex coefficients α and β. The time constant associated with the temporal decay of this phase relation is the T_2 time. In fact, we can associate a density matrix with the qubit (recall Chapter 5 material):

$$\rho_{qubit} = \begin{pmatrix} |\alpha|^2 & \alpha^*\beta \\ \beta^*\alpha & |\beta|^2 \end{pmatrix} . \tag{8.31}$$

The off-diagonal terms represent the phase relationship between α and β. The quantity T_2 is the time constant associated with the temporal decay of the off-diagonal terms, and therefore provides a measure of the robustness of a qubit.

Whether T_1 will be larger than T_2, or the other way around, depends on the primary mode of spin relaxation. Golovach et al. [29] have argued that if the spin relaxation is mediated by Dresselhaus and Rashba spin-orbit interaction, then $T_2 = 2T_1$. However, for other modes of spin relaxation, it is very possible that $T_2 < T_1$ (see Equation (8.32) later).

It is difficult to measure the T_2 time directly since we have to probe decoherence of a single electron. In most experiments, one probes a large number of spins simultaneously. Even if these spins do not interact with each other, inhomogeneities in local magnetic fields caused by nuclear spins having random polarization, or spatially varying spin-orbit interaction, will increase the decoherence rate substantially. Therefore, what we measure will always be less than the actual T_2 time associated with spin-spin interaction encountered by a single electron. We will call the measured time the T_2^* time. It has been shown that T_2^* can be several times smaller than T_2 [30].

There are many ways of measuring T_2^*. One of the most direct methods is *electron spin resonance* (ESR). In ESR experiments, we apply a dc magnetic field to lift the degeneracy between up-spin and down-spin energies via the Zeeman effect. An ac magnetic field, produced by a microwave source, is then turned on. It induces Rabi oscillations between the Zeeman-split levels when the microwave frequency is resonant with the spin-splitting energy (recall Chapter 4 material). Consequently, microwave is absorbed when resonance occurs. The Rabi oscillation amplitude decays in time roughly exponentially, with a time constant T_2^*. The Fourier transform of the damped oscillation is the absorption spectrum (absorption intensity versus microwave frequency). The linewidth of the absorption peak (which should have a Lorentzian shape) is $1/T_2^*$. Consequently, from the absorption peak linewidth, we can deduce the T_2^* time. Although this is a straightforward technique, it has been seldom applied to quantum dots since the absorption intensity depends on the number of participating spins and that number has to be large in order to produce a strong enough signal. With quantum dots, the number of spins involved is

typically small so that normally ESR is not a reliable method. However, if the density of quantum dots is sufficiently high, then it may be possible to obtain a strong enough ESR signal (particularly at low temperatures) to yield a reliable estimate of the T_2^* time. These days, it is possible to produce extremely dense array of quantum dots (density $> 10^{11}$ cm^{-2}) using electrochemical self assembly [31]. With such dense arrays, ESR can be applied successfully to measure the T_2^* time.

There are several reports of the measurement of T_1 times in quantum dots [21, 28, 24, 25], but few reports of the measurement of T_2 times (single spin decoherence time). That would require a specific $\pi/2$-π spin echo sequence [32], which is difficult to implement. Such measurements have been carried out in some systems, notably nitrogen vacancy color centers in diamond. The measured T_2 time of spin dephasing in the color center was amazingly long – 58 μs at room temperature (300 K) [33]. Coherent coupling between single spins in diamond was demonstrated at room temperature by Gaebel, et al. [34], while Dutt and co-workers have now demonstrated a quantum register based on individual electronic and nuclear spin-based qubits in this system [35].

The *ensemble averaged* T_2^* time (decoherence time associated with many spins) has been measured in 2-8 nm CdSe quantum dots using femtosecond resolved Faraday rotation [36]. Dzhioev, et al. have measured the T_2^* time associated with the decoherence of several interacting spins in an n-GaAs heterostructure using polarized photoluminescence measurement [37]. They found that the T_2^* time is determined by hyperfine interaction with nuclear spins and is of the order of few hundred nanoseconds. In our laboratory, we too have measured the T_2^* time in electrochemically self assembled cylindrical CdS quantum dots, of 10 nm diameter, using direct ESR spectroscopy [38]. The measured T_2^* time was a few nanoseconds. Surprisingly, we found that the T_2^* time increases with temperature. This can be explained by invoking dephasing due to hyperfine interaction with nuclear spins.

It is widely believed that the primary cause of spin dephasing in an ensemble of quantum dots is the spatially varying hyperfine interaction caused by nuclear spins [18, 30, 39, 40]. The spatial variation comes about from two sources: First, the electron wavefunction varies in space and the strength of the hyperfine interaction depends on this wavefunction. Second, the magnetic field due to nuclear spins is also inhomogeneous. This inhomogeneity is the major contributor to the dephasing rate $1/T_2^*$. However, it is possible in singly charged quantum dots to synchronize the electron spin precession about a magnetic field into well-defined modes with a laser pulse sequence. In that case, the light stimulated fluctuations of the hyperfine nuclear field acting on the electron are suppressed. The information about the electron spin precession can be imprinted on the nuclei and stored for tens of minutes in darkness at a temperature of 6 K. This was recently demonstrated by Greilich, et al [41].

In quantum wells, on the other hand, the major spin dephasing mechanism

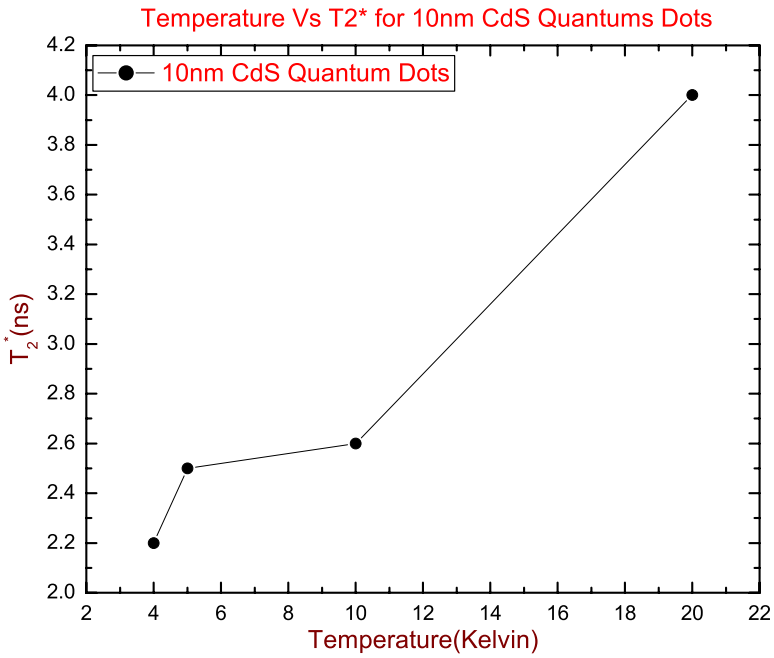


FIGURE 8.5
Measured T_2^* time as a function of temperature in CdS quantum dots of 50-nm diameter. Reproduced from B. Kanchibotla, et al., Proceedings of IEEE NANO 2006, Cincinnati, 2006, with permission from the Institute of Electrical and Electronics Engineers (copyright Institute of Electrical and Electronics Engineers). The T_2 time increases with temperature since the primary relaxation mechanism is hyperfine interaction with nuclear spins. The strength of this interaction decreases with increasing temperature because of increasing depolarization of nuclear spins.

is believed to be Rashba spin-orbit interaction [42]. That interaction, however, is extremely weak in quantum dots since the electron motion is restricted in all three dimensions [43].

Reference [18] gives an equation for the transverse relaxation rate due to hyperfine interactions with nuclear spins in the special case of electrons hopping between dots:

$$\frac{1}{T_2} = \frac{1}{2T_1} + \frac{(g\mu_B)^2}{2\hbar^2} \int_{-\infty}^{\infty} d\tau < H_{NZ}(0) H_{NZ}(\tau) >, \tag{8.32}$$

where the bracket $< .. >$ denotes ensemble average and H_{NZ} is the effective magnetic field due to nuclear spins. Obviously, the dephasing rate increases with increasing autocorrelation given by the integral in Equation (8.32). With increasing temperature, the nuclear spin polarizations are randomized leading to a decrease in the autocorrelation function. That explains why the dephasing rate decreases with increasing temperature in the experiment of ref. [38].

The thermal depolarization also plays a direct role as mentioned in ref. [40]. For an unpolarized configuration of nuclear spins, the net nuclear magnetic field is $A/(\sqrt{N} g\mu_B)$ (where A is the hyperfine constant and N is the number of nuclei in a quantum dot, typically about 10^5). For a fully polarized configuration, the net field is $A/(g\mu_B)$. Therefore, thermally induced depolarization of nuclear spins will tend to reduce the dephasing rate by reducing the net nuclear magnetic field by a factor of $\sqrt{N} \sim 300$ in typical quantum dots.

Using optical pump and probe techniques, spin state of carriers and nuclei in *individual* semiconductor quantum dots have been studied experimentally [44, 45]. These measurements have revealed that the primary decoherence mechanism limiting T_2 is hyperfine interaction of the conduction electron spins with nuclear spins. The measured T_2 time in individual GaAs quantum dots is of the order of 10 nanoseconds, which is $\hbar\sqrt{N}/A$. Thus, the experimentally measured T_2 times in semiconductor dots are much smaller (by about 6 orders of magnitude) than the T_1 times, which clearly demonstrates that realizing *quantum* logic processors based on single electron spins in quantum dots is a much more difficult challenge than realizing *classical* logic processors (see Chapter 13). The latter does not require phase coherence and merely requires a long T_1 time, whereas the former requires phase coherence and therefore a long T_2 time, as well as a long T_1 time.

8.4 The Spin Galvanic Effect

In a spin polarized two-dimensional electron gas, spin relaxing scattering can cause strange effects in the presence of spin-orbit interaction. One example is the *Spin Galvanic Effect*, where an electric current (charge current) can flow

without a battery! [46]. This phenomenon is explained below, following ref. [46].

Electric current is usually generated by (i) electric or magnetic fields (drift current), or (ii) a spatial gradient of carrier concentration (diffusion current) or (iii) temperature (thermogalvanic current). In a spin-polarized electron gas, an electric current can be generated without any electric/magnetic fields, or without a concentration or thermal gradient. This is the Spin Galvanic Effect.

Consider a two-dimensional electron gas in the x-z plane. The energy dispersion relation (E versus k_x for a fixed k_z) is shown in Fig. 8.6. For the sake of simplicity, let us assume that there is only Rashba or Dresselhaus interaction (but not both) so that each band has a fixed spin quantization axis. The E-k plots are two displaced parabolas as shown in Fig. 8.6. Since the electron gas is spin polarized, we show that one band is filled more than the other, meaning that it is occupied to higher energy than the other.

As long as the carrier distribution in either $E - k$ parabola is symmetric about the corresponding band minima at $k = k_{x1,x2}$, there can be no net current flow since for every carrier with a velocity v_x, there is a carrier with velocity -v_x. However, asymmetric k-dependent scattering can occur that will cause an imbalance in carrier population, making the population of carriers with positive v_x slightly exceed that with negative v_x, or vice versa. In that case, a charge current flows without any electric or magnetic field, or without any concentration or thermal gradient. There is no violation of energy conservation since energy must be constantly supplied to maintain the non-equilibrium distribution of spin polarized carriers. Typically, the spin polarized carrier population is created by optical orientation using circularly polarized light [47].

Because of spin-orbit interaction, the energy dispersion parabolas are horizontally shifted from each other. As a result, the wavevector transfer Δk_x associated with a scattering event that flips a left traveling down-spin electron in one band to a right traveling up-spin electron in the other band (with a change in the direction of velocity) is different from that associated with the scattering event that flips a right traveling down-spin electron in the first band to left traveling up-spin electron in the second band. In other words, the wavevector transfer Δk_x depends on the *initial direction of velocity*. Now, the spin flip scattering could be caused by a charged magnetic impurity that interacts with an electron via Coulomb interaction. The probability of that scattering is proportional to $1/(\Delta k_x)^2$, if we neglect screening. Therefore, the probability of spin flip scattering, with an associated change in the direction of velocity, depends on the *sign* of the initial velocity. In the example shown in Fig. 8.6, there is a higher probability of a right traveling electron flipping to a left traveling electron. Therefore, the population of left traveling electrons will increase at the expense of right traveling electrons, causing an imbalance that results in a net current flow without a battery, concentration gradient or thermal gradient. This is the basis of the Spin Galvanic Effect.

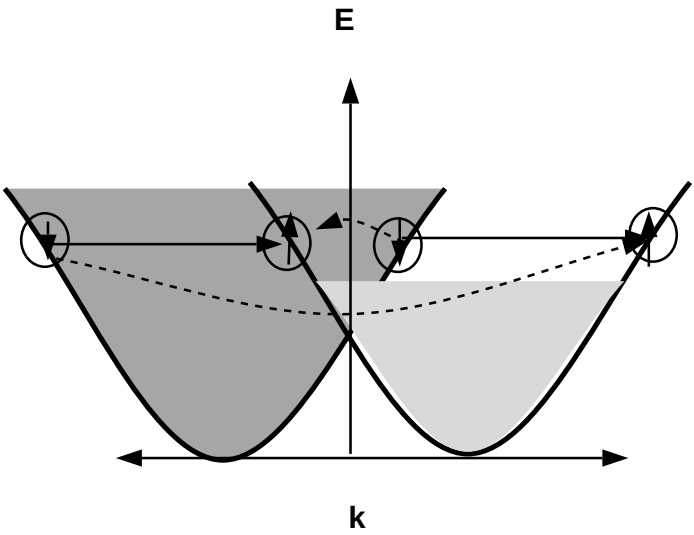


FIGURE 8.6
Understanding the Spin Galvanic Effect. The energy dispersion relation in the presence of Rashba interaction in a spin polarized electron gas. We assume that the down-spin band has more carriers than the up-spin band, so that the down-spin band is occupied to higher energy. Pauli exclusion principle will allow an electron from a filled state in the down-spin band to scatter to an empty state in the up-spin band. The spin relaxing events are designated with arrows. We show an elastic scattering event, but the arguments presented here hold for inelastic events as well. Because of spin orbit interaction, the energy dispersion parabolas are horizontally shifted from each other. As a result, the wavevector transfer associated with scattering events that result in a change in the direction of velocity (shown with broken arrows) depends on the sign of the initial velocity. This gives rise to the Spin Galvanic Effect. Adapted from ref. [46], Reprinted by permission from Macmillan Publishers Ltd: [Nature (London), **417**, 153, copyright 2002].

8.5 Problems

- **Problem 8.1**

  Consider an electron with x-polarized spin placed in a magnetic field directed along the z-coordinate. Show using quantum mechanics that if there are no dissipative processes, then the spin will precess about the magnetic field according to Equation (8.6).

  Solution

  First note that the x-polarized spin is not parallel or anti-parallel to the z-directed magnetic field. Therefore at time $t = 0$, the spin is *not* in an eigenstate of the Hamiltonian containing the z-directed magnetic field. Consequently, the spin polarization must change with time and we have to show that the spin will precess about the magnetic field.

  If there are no dissipative processes, then the eigenspinor $[\phi]$ will obey the *time dependent* Pauli Equation

$$\left[i\hbar\frac{\partial}{\partial t} + < H_0 > + \frac{g}{2}\mu_B \vec{B}\cdot\vec{\sigma}\right][\phi] = 0, \tag{8.33}$$

  which has the solution

$$\begin{aligned}[\phi](t) &= exp[-i(< H_0 > + (g/2)\mu_B \vec{B}\cdot\vec{\sigma})t/\hbar][\phi](0) \\ &= \begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}^{-1}\begin{bmatrix}e^{-iE_+ t/\hbar} & 0\\ 0 & e^{-iE_- t/\hbar}\end{bmatrix}\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}[\phi](0),\end{aligned} \tag{8.34}$$

  where E_- and E_+ are the eigenvalues of the 2×2 matrix $\left[< H_0 > + \frac{g}{2}\mu_B \vec{B}\cdot\vec{\sigma}\right]$. Since the magnetic field is in the z-direction, the energy eigenstates and eigenspinors of $\left[< H_0 > + \frac{g}{2}\mu_B \vec{B}\cdot\vec{\sigma}\right]$ will be given by Equations (8.3) and (8.5).

  The last equation tells us what the eigenspinor will be at any instant of time t.

  Since the initial spin polarization was along the x-direction, therefore $[\phi](0)$ is the x-polarized state $(1/\sqrt{2})[1 \quad -1]^\dagger$. As a result, the last equation reduces to

$$[\phi](t) = \begin{bmatrix} e^{-iE_+ t/\hbar} \\ -e^{-iE_- t/\hbar}\end{bmatrix}. \tag{8.35}$$

  Now the spin components will be given by

$$\begin{aligned}S_x(t) &= [\phi]^\dagger(t)[\sigma_x][\phi]^\dagger(t) \\ &= \left[e^{+iE_+ t/\hbar}/\sqrt{2} \; -e^{+iE_- t/\hbar}/\sqrt{2}\right]\begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix}\begin{bmatrix} e^{-iE_+ t/\hbar}/\sqrt{2} \\ -e^{-iE_- t/\hbar}/\sqrt{2}\end{bmatrix} \\ &= -cos\theta,\end{aligned} \tag{8.36}$$

where $\theta = (E_+ - E_-)t/\hbar = g\mu_B Bt/\hbar$ where B is the flux density associated with the magnetic field.

Similarly,

$$\begin{aligned} S_y(t) &= [\phi]^\dagger(t)[\sigma_y][\phi]^\dagger(t) \\ &= \left[e^{+iE_+t/\hbar}/\sqrt{2} \; -e^{+iE_-t/\hbar}/\sqrt{2} \right] \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix} \begin{bmatrix} e^{-iE_+t/\hbar}/\sqrt{2} \\ -e^{-iE_-t/\hbar}/\sqrt{2} \end{bmatrix} \\ &= -sin\theta \end{aligned} \tag{8.37}$$

and

$$\begin{aligned} S_z(t) &= [\phi]^\dagger(t)[\sigma_y][\phi]^\dagger(t) \\ &= \left[e^{+iE_+t/\hbar}/\sqrt{2} \; -e^{+iE_-t/\hbar}/\sqrt{2} \right] \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} e^{-iE_+t/\hbar}/\sqrt{2} \\ -e^{-iE_-t/\hbar}/\sqrt{2} \end{bmatrix} \\ &= 0. \end{aligned} \tag{8.38}$$

Now

$$\begin{aligned} \frac{d\vec{S}}{dt} &= \hat{x}\frac{dS_x}{dt} + \hat{y}\frac{dS_y}{dt} + \hat{z}\frac{dS_z}{dt} \\ &= -\hat{x}\frac{d(cos\theta)}{dt} + -\hat{y}\frac{d(sin\theta)}{dt} \\ &= \frac{d\theta}{dt}\left[\hat{x}sin\theta - \hat{y}cos\theta\right] \\ &= \frac{g}{2}\frac{eB}{m}\left[\hat{x}sin\theta - \hat{y}cos\theta\right] \\ &= \frac{eB}{m*}\left[\hat{x}sin\theta - \hat{y}cos\theta\right], \end{aligned} \tag{8.39}$$

where we have used the fact that normally $g = 2m/m^*$ with m being the free electron mass and m^* the effective mass of the electron.

Now, since the B-field is directed along the z-direction,

$$\begin{aligned} \vec{\Omega} \times \vec{S} &= \frac{e\vec{B}}{m^*} \times \vec{S} \\ &= \frac{eB}{m^*}\left[-\hat{x}S_y + \hat{y}S_x\right] \\ &= \frac{eB}{m^*}\left[\hat{x}sin\theta - \hat{y}cos\theta\right]. \end{aligned} \tag{8.40}$$

The last two equations show that

$$\frac{d\vec{S}}{dt} = \vec{\Omega} \times \vec{S}, \tag{8.41}$$

which is the result to be proved.

- **Problem 8.2**

  Derive Equation (8.18).

  Hint: Set the expectation values of the Rashba and Dresselhaus Hamiltonians in Equations (7.1) and (7.2) [for a one-dimensional system with axis along the x-direction] equal to the effective Zeeman splitting energies $g\mu_B B_{Rashba}$ and $g\mu_B B_{Dresselhaus}$ and replace $\hbar k_x$ with $m^* v_x$.

- **Problem 8.3**

  Show that in a quantum wire, three ingredients must be simultaneously present to cause D'yakonov-Perel' spin relaxation in *space*: (a) Rashba interaction, (b) Dresselhaus interaction and (c) multi-channeled transport. How is the situation different for spin relaxation in *time*?

  Solution

  Consider the quantum wire of Fig. 8.4. We will assume that electrons are injected from the left (at $x = 0$) with their spins polarized along the wire axis, so that

$$\begin{aligned} S_x &= 1 \\ S_y &= 0 \\ S_z &= 0. \end{aligned} \tag{8.42}$$

  In this system, the spin precession frequency $\vec{\Omega}(\vec{v})$ is given by

$$\vec{\Omega}(\vec{v}) = \Omega_D(\vec{v})\hat{x} + \Omega_R(\vec{v})\hat{z}, \tag{8.43}$$

  where $\hat{n}$ is the unit vector along the n-axis. In the above, the first term is the contribution due to the Dresselhaus interaction and the second term is the contribution due to Rashba interaction (recall Equation(8.18)). These quantities are given by

$$\begin{aligned} \Omega_D(\vec{v}) &= \frac{2a_{42}m^*}{\hbar^2}\left[\left(\frac{m\pi}{W_z}\right)^2 - \left(\frac{n\pi}{W_y}\right)^2\right] v = \Omega_{D0} v \\ \Omega_R(\vec{v}) &= \frac{2a_{46}m^*}{\hbar^2} E_y v = \Omega_{R0} v. \end{aligned} \tag{8.44}$$

  The spin precesses according to Equation (8.6):

$$\begin{aligned} \frac{d\vec{S}}{dt} &= \hat{x}\frac{dS_x}{dt} + \hat{y}\frac{dS_y}{dt} + \hat{z}\frac{dS_z}{dt} \\ &= \vec{\Omega}(\vec{v}) \times \vec{S} \\ &= det \begin{bmatrix} \hat{x} & \hat{y} & \hat{z} \\ \Omega_D(\vec{v}) & 0 & \Omega_R(\vec{v}) \\ S_x & S_y & s_z \end{bmatrix} \\ &= -\hat{x}\left(\Omega_R(\vec{v})S_y\right) - \hat{y}\left(\Omega_D(\vec{v})S_z - \Omega_R(\vec{v})S_x\right) + \hat{z}\left(\Omega_D(\vec{v})S_y\right) \ , \end{aligned} \tag{8.45}$$

where S_n is the spin component along the n-axis.

Equating each component separately and using the chain rule of differentiation, we get:

$$\begin{aligned}
\frac{dS_x}{dx}\frac{dx}{dt} &= \frac{dS_x}{dx}v = -\Omega_{R0}vS_y \\
\frac{dS_y}{dx}\frac{dx}{dt} &= \frac{dS_y}{dx}v = \Omega_{R0}vS_x - \Omega_{D0}vS_z \\
\frac{dS_z}{dx}\frac{dx}{dt} &= \frac{dS_z}{dx}v = \Omega_{D0}vS_y.
\end{aligned} \tag{8.46}$$

First, let us derive an expression for S_y as a function of position x. From the second equation above, we get by further differentiation

$$\begin{aligned}
\frac{d^2S_y}{dx^2} &= \Omega_{R0}\frac{dS_x}{dx} - \Omega_{D0}\frac{dS_z}{dx} \\
&= -\Omega_{R0}\left(\Omega_{R0}S_y\right) - \Omega_{D0}\left(\Omega_{D0}S_y\right) \\
&= -\left(\Omega_{R0}^2 + \Omega_{D0}^2\right)S_y \\
&= \Omega_0^2 S_y,
\end{aligned} \tag{8.47}$$

where $\Omega_0^2 = \Omega_{R0}^2 + \Omega_{D0}^2$.

The solution of the above equation is

$$S_y(x) = Acos\left(\Omega_0 x\right) + Bsin\left(\Omega_0 x\right). \tag{8.48}$$

Using the boundary condition in Equation (8.42), we reduce the above to

$$S_y(x) = Bsin\left(\Omega_0 x\right). \tag{8.49}$$

Differentiating the above and using Equation (8.46), we get

$$\frac{dS_y}{dx} = B\Omega_0 cos\left(\Omega_0 x\right) = \Omega_{R0}S_x - \Omega_{D0}S_z. \tag{8.50}$$

Again, using the boundary condition at $x = 0$ from Equation (8.42) in the above, we get that $B = \Omega_{R0}/\Omega_0$, so that

$$S_y(x) = \frac{\Omega_{R0}}{\Omega_0}sin\left(\Omega_0 x\right). \tag{8.51}$$

Next, let us derive an expression for S_x as a function of x. From Equation (8.46) and the above solution for S_y, we get

$$\frac{dS_x}{dx} = -\Omega_{R0}S_y = -\frac{\Omega_{R0}^2}{\Omega_0}sin\left(\Omega_0 x\right). \tag{8.52}$$

Integration of the above yields

$$S_x(x) = \frac{\Omega_{R0}^2}{\Omega_0^2} cos\left(\Omega_0 x\right) + C, \tag{8.53}$$

where C is a constant to be determined from the boundary condition in Equation (8.42). This determination yields $C = 1 - \Omega_{R0}^2/\Omega_0^2 = \Omega_{D0}^2/\Omega_0^2$. Therefore, the expression for S_x becomes

$$S_x(x) = \frac{\Omega_{R0}^2}{\Omega_0^2} cos\left(\Omega_0 x\right) + \frac{\Omega_{D0}^2}{\Omega_0^2} = 1 - 2\frac{\Omega_{R0}^2}{\Omega_0^2} sin^2\left(\frac{\Omega_0 x}{2}\right). \tag{8.54}$$

Finally, we derive an expression for S_z. Using Equation (8.46) and the expression for S_y, we get

$$\frac{dS_z}{dx} = \Omega_{D0} S_y = \frac{\Omega_{R0}\Omega_{D0}}{\Omega_0} sin\left(\Omega_0 x\right). \tag{8.55}$$

Integrating the above, we obtain:

$$S_z(x) = -\frac{\Omega_{R0}\Omega_{D0}}{\Omega_0^2} cos\left(\Omega_0 x\right) + D, \tag{8.56}$$

where D is a constant determined from boundary condition (Equation (8.42)) to be $\Omega_{R0}\Omega_{D0}/\Omega_0^2$. Therefore,

$$S_z(x) = \frac{\Omega_{R0}\Omega_{D0}}{\Omega_0^2}\left[1 - cos\left(\Omega_0 x\right)\right] = 2\frac{\Omega_{R0}\Omega_{D0}}{\Omega_0^2} sin^2\left(\frac{\Omega_0 x}{2}\right). \tag{8.57}$$

Note from Equations (8.54), (8.51) and (8.57), that the spin components at any position in space are *independent* of the carrier velocity v since the velocity "canceled out" in Equation (8.46). Therefore carrier scattering, which changes v, has no direct effect on the spin components. It may have an indirect effect through Ω_{R0} and Ω_{D0}, if these quantities are affected by scattering. Ω_{R0} certainly is not affected and Ω_{D0} is not either unless there is inter-subband scattering that changes the subband indices m and n in Equation (8.44) and causes multi-channeled transport where more than one subband will carry current. Therefore, we conclude that *intra-subband scattering does not affect the spin components in space.* This is a remarkable result, which leads to the absence of D-P spin relaxation in space as long as inter-subband scattering (leading to multi-channeled transport) is absent. We point out that this happens because the spin precession frequencies Ω_D and Ω_R, **which are spin-independent**, are each proportional to the spin-independent magnetic field and therefore proportional to the group velocity v and *not* wavevector k. If Ω_D and Ω_R were proportional to k instead of v, then v would

not have canceled out in Equation (8.46) and the spin components at any coordinate position in space would not have been unaffected by scattering. As a result, there would have been D-P relaxation in space.

Examining the expressions for $S_x(x)$, $S_y(x)$ and $S_z(x)$ given by Equations (8.51), (8.54) and (8.57), we see immediately that if Rashba interaction is absent ($\Omega_{R0} = 0$), then $S_x(x) = 1$, $S_y(x) = 0$ and $S_z(x) = 0$, even if there is inter-subband scattering. Therefore, the original spin orientation does not change in space at all and there is no D-P spin relaxation. Of course, it is not necessary that the original spin orientation remain intact for D-P relaxation to not occur. What is necessary is that even if the orientation changes in space, it must change in exactly the same way for every electron so that the ensemble averaged magnitude of the spin $| < S > |$ does not decay in space (here the angular bracket $< .. >$ denotes ensemble averaging). This will happen if $| < S > |^2 = < S_x > (x)^2 + < S_y > (x)^2 + < S_z > (x)^2$ is independent of x.

Next we consider what happens if the Dresselhaus interaction is absent. In that case, $S_x(x) = cos\,(\Omega_0 x)$, $S_y(x) = sin\,(\Omega_0 x)$ and $S_z(x) = 0$. Here, the spin orientation does change in space, but the quantity $|S|^2 = S_x(x)^2 + S_y(x)^2 + S_z(x)^2 = cos^2\,(\Omega_0 x) + sin^2\,(\Omega_0 x) = 1$, independent of x. Therefore, there is no D-P spin relaxation, even if there were inter-subband scattering. Inter-subband scattering is relevant to Dresselhaus interaction only; therefore if the Dresselhaus interaction is absent, inter-subband scattering is immaterial.

Now let us consider the situation when both Dresselhaus and Rashba interaction are present. As long as transport is single channeled and there is no inter-subband scattering, Ω_{D0} and Ω_0 are the same for every electron. In that case, it is easy to show that $< S_x > (x) = S_x(x)$, $< S_y > (x) = S_y(x)$ and $< S_z > (x) = S_z(x)$.

Therefore,

$$
\begin{aligned}
| < S > |^2 &= < S_x > (x)^2 + < S_y > (x)^2 + < S_z > (x)^2 \\
&= S_x(x)^2 + S_y(x)^2 + S_z(x)^2 \\
&= 1 + 4\left(\frac{\Omega_{R0}}{\Omega_0}\right)^4 sin^4\left(\frac{\Omega_0 x}{2}\right) - 4\left(\frac{\Omega_{R0}}{\Omega_0}\right)^2 sin^2\left(\frac{\Omega_0 x}{2}\right) \\
&\quad + 4\left(\frac{\Omega_{R0}}{\Omega_0}\right)^2 sin^2\left(\frac{\Omega_0 x}{2}\right) cos^2\left(\frac{\Omega_0 x}{2}\right) \\
&\quad + 4\left(\frac{\sqrt{\Omega_{R0}\Omega_{D0}}}{\Omega_0}\right)^4 sin^4\left(\frac{\Omega_0 x}{2}\right) \\
&= 1 + 4\left(\frac{\Omega_{R0}}{\Omega_0}\right)^4 sin^4\left(\frac{\Omega_0 x}{2}\right) - 4\left(\frac{\Omega_{R0}}{\Omega_0}\right)^2 sin^4\left(\frac{\Omega_0 x}{2}\right) \\
&\quad + 4\frac{\Omega_{R0}^2(\Omega_0^2 - \Omega_{R0}^2)}{\Omega_0^4} sin^4\left(\frac{\Omega_0 x}{2}\right)
\end{aligned}
$$

$$= 1. \tag{8.58}$$

Thus, we have shown that as long as transport is single channeled, there is no D-P relaxation. Combining all of the above, we conclude that in a quantum wire, three ingredients must be simultaneously present to cause D-P spin relaxation in space: Rashba interaction, Dresselhaus interaction and multi-channeled transport. In our proof, we assume that the initial spin polarization at $x = 0$ is along the wire axis. The reader can easily repeat the proof if the initial polarization was along any of the other two directions – y or z. A slightly different (more involved) proof of this result can be found in ref. [10].

We now revisit the issue of whether there is D-P relaxation in time. The difference between *time* and *space* is the following: in the case of "space", we ask what are the spin components $S_x(x)$, $S_y(x)$ and $S_z(x)$ at a position x, irrespective of how long it took the electron to get there. The electron may have visited the coordinate x at some earlier time, then gone forward under the action of an accelerating electric field, subsequently backscattered to a position behind x, and again gone forward under the electric field to reach the position x a second time. What we have shown here is that the spin components are identical during the first and second visit to the location x! In other words, the past history does not matter. An electron may visit the location x many times, and each time it will have exactly the same spin components at x as long as we have single channeled transport. No amount of intra-subband scattering can change this result! This is certainly counter-intuitive. Not only that, but every electron at positions x has exactly the same spin components, even though they may have had very different scattering histories! That too is counter-intuitive.

In the case of "time", we ask what are the spin components of different electrons, irrespective of what their locations are in the quantum wire, if we take a snapshot at time t. We know that these spin components will be different, depending on the location, since we have derived the spin components as a function of x. Thus, if we ensemble average over all electrons, this average will change with time t. The change may not be a monotonic decay with time. There may be oscillations and non-monotonic behavior. But nontheless there is a temporal change and that can be interpreted as "spin relaxation".

The interesting feature is that there is distinct difference between relaxation in *space* and relaxation in *time*. The relaxation length in space is infinity if transport is single channeled, but the relaxation time is finite. Drift-diffusion models of spin transport often designate the relaxation length as the spin diffusion length $L_s = \sqrt{D\tau_s}$ where D is the spin diffusion coefficient and τ_s is the spin relaxation time. We are now faced with the following dilemma: $L_s \to \infty$, but τ_s is finite. This can be

resolved in only two ways: (i) $D_s \rightarrow \infty$, or (ii) the drift-diffusion model that stipulates $L_s = \sqrt{D\tau_s}$ is completely invalid. We will address the shortcomings of drift-diffusion models in Chapter 9, but for the time being let us ponder the possibility that $D_s \rightarrow \infty$. We can be certain that the "charge diffusion coefficient" D_c is finite, since there is plenty of scattering. Therefore, if the drift-diffusion model remains valid, we must conclude that $D_c \neq D_s$. In other words, charge and spin diffusion coefficients are very different. There seems to be scant appreciation of this fact within the spintronics community. Most researchers assume tacitly (and incorrectly) that the two diffusion coefficients are always the same. In the case of Elliott-Yafet relaxation, they are probably not too different since momentum relaxing scattering, which contributes to charge diffusion coefficient, is also the cause of spin relaxation. Moreover, the charge and spin relaxation rates are proportional to each other. However, in the case of D'yakonov-Perel' relaxation, the two rates are *inversely* proportional to each other. There is no reason to assume *a priori* that if the D'yakonov-Perel' mechanism is the major cause of spin relaxation, the two diffusion coefficients will be equal.

- **Problem 8.4**

  Electrons are injected into a quantum wire with a constant velocity. There is a constant electric field driving transport and there are no momentum randomizing collisions (transport is ballistic). A single subband is occupied. Show that there is no D'yakonov-Perel' relaxation in *time*.

  Solution

  Let the initial injection velocity $= u$. The velocity at any time t is then given by

$$v(t) = u + ft, \tag{8.59}$$

  where f is the acceleration $= \mathcal{E}/m^*$ with $\mathcal{E}$ being the electric field driving transport. The above equation is valid *for every electron* in the ensemble since there is no scattering to change the velocities of different electrons randomly.

  From Equation (8.46)

$$\begin{aligned} \frac{dS_x}{dt} &= -\Omega_{R0} v(t) S_y \\ \frac{dS_y}{dt} &= \Omega_{R0} v(t) S_x - \Omega_{D0} v(t) S_z \\ \frac{dS_z}{dt} &= \Omega_{D0} v(t) S_y. \end{aligned} \tag{8.60}$$

  Since $v(t)$ is the same for every electron, we can replace the spin component S_n in the above equation with its ensemble averaged value $< S_n >$.

Therefore,

$$\begin{aligned}
\frac{d|<S>|^2}{dt} &= \frac{d<S_x>^2}{dt} + \frac{d<S_y>^2}{dt} + \frac{d<S_z>^2}{dt} \\
&= 2<S_x>\frac{d<S_x>}{dt} + 2<S_y>\frac{d<S_y>}{dt} \\
&\quad +2<S_z>\frac{d<S_z>}{dt} \\
&= -2\Omega_{R0}v(t)<S_y><S_x> +2\Omega_{R0}v(t)<S_x><S_y> \\
&\quad -2\Omega_{D0}v(t)<S_z><S_y> +2\Omega_{D0}v(t)<S_y><S_z> \\
&\quad -2\Omega_{R0}f<S_y><S_x> +2\Omega_{R0}f<S_x><S_y> \\
&\quad -2\Omega_{D0}f<S_z><S_y> +2\Omega_{D0}f<S_y><S_z> \\
&= 0. \qquad (8.61)
\end{aligned}$$

Therefore, $|<S>|$ will not decay in time and there will be no D'yakonov-Perel' spin relaxation in time. This exercise will show the reader that there is a simpler solution to Prob. 8.3, but we presented a more involved solution since we wanted to derive expressions for the individual spin components S_n.

- **Problem 8.5**

  If the spin-orbit magnetic field B_{so} were proportional to wavevector k_x and not velocity v_x, then you can show easily that there will be D'yakonov-Perel' spin relaxation in space in a quantum wire with a single subband occupied. Use Equation (7.53) to relate v_x to k_x and derive this result. This result is of course unphysical since there should never be any D'yakonov-Perel' relaxation in strictly single-channeled transport. Does this mean that the spin-independent field gives the right result and the spin-dependent field does not? Not necessarily. In the analysis you would have assumed a spin-independent precession frequency Ω which is proportional to B_{so}. Therefore, B_{so} should also be spin-independent and you cannot use the spin-dependent B_{so} which is proportional to k.

8.6 References

[1] J. Fabian and S. Das Sarma, "Spin relaxation of conduction electrons", J. Vac. Sci. Technol. B, **17**, 1708 (1999).

[2] R. J. Elliott, "Theory of the effect of spin orbit coupling on magnetic resonances in some semiconductors", Phys. Rev., **96**, 266 (1954).

[3] G. E. Pikus and A. N. Titkov, in *Optical Orientation*, Eds. F. Meier and B. P. Zakharchenya (North Holland, Amsterdam, 1984).

[4] F. J. Jedema, M. S. Nijboer, A. T. Philip and B. J. Van Wees, "Spin injection and spin accumulation in all metal spin valves", Phys. Rev. B, **67**, 085319 (2003).

[5] Y. Yafet in *Solid State Physics*, Eds. F. Seitz and D. Turnbull, Vol. 14 (Academic, New York, 1963).

[6] F. Monod and P. Beunue, "Conduction electron spin flip by phonons in metals: Analysis of experimental data", Phys. Rev. B, **19**, 911 (1979).

[7] S. Pramanik, S. Bandyopadhyay and M. Cahay, "Spin relaxation of "upstream" electrons in quantum wires: Failure of the drift diffusion model", Phys. Rev. B, **73**, 125309 (2006).

[8] J. S. Sandhu, A. P. Heberle, J. J. Baumberg and J. R. A. Cleaver, "Gateable suppression of spin relaxation in semiconductors", Phys. Rev. Lett., **86**, 2150 (2001).

[9] J. Kainz, U. Rössler and R. Winkler, "Temperature dependence of D'yakonov-Perel' spin relaxation in zinc blende semiconductor quantum structures", Phys. Rev. B, **70**, 195322 (2004).

[10] S. Pramanik, S. Bandyopadhyay and M. Cahay, "Spin relaxation in the channel of a Spin Field Effect Transistor", IEEE Trans. Nanotech., **4**, 2 (2005).

[11] A. G. Mal'shukov and K. A. Chao, "Waveguide diffusion modes and slowdown of D'yakonov-Perel' spin relaxation in narrow two dimensional semiconductor channels", Phys. Rev. B, **61**, R2413 (2000).

[12] S. Pramanik, S. Bandyopadhyay and M. Cahay, "Decay of spin polarized hot carrier current in a quasi one-dimensional spin valve structure", Appl. Phys. Lett., **84**, 266 (2004).

[13] A. W. Holleitner, V. Sih, R. C. Myers, A. C. Gossard and D. D. Awschalom, "Suppression of spin relaxation in submicron InGaAs wires", www.arXiv.org/cond-mat/0602155.

[14] S. Pramanik, S. Bandyopadhyay and M. Cahay, "Spin dephasing in quantum wires", Phys. Rev. B, **68**, 075313 (2003).

[15] T. P. Pareek and P. Bruno, "Spin coherence in two dimensional electron gas with Rashba spin orbit interaction", Phys. Rev. B, **65**, 241305(R) (2002).

[16] N. S. Averkiev and L. E. Golub, "Giant spin relaxation anisotropy in zinc blende heterostructures", Phys. Rev. B, **60**, 15582 (1999).

[17] Y. Ohno, R. Terauchi, T. Adachi, F. Matsukura and H. Ohno, "Spin relaxation in GaAs(110) quantum wells", Phys. Rev. Lett., **83**, 4196 (1999).

[18] A. V. Khaetski, D. Loss and L. Glazman, "Electron spin evolution induced by interaction with nuclei in a quantum dot", Phys. Rev. B, **67**, 195329 (2003).

[19] A. V. Khaetski and Y. V. Nazarov, "Spin flip transitions between Zeeman sublevels in semiconductor quantum dots", Phys. Rev. B, **61**, 12639 (2000).

[20] S. Amasha, K. MacLean, Iuliana Radu, D. M. Zümbuhl, M. A. Kastner, M. P. Hanson and A. C. Gossard, "Electrical control of spin relaxation in a quantum dot", arXiv:0707.1656.

[21] R. Hanson, et al., "Zeeman energy and spin relaxation in one electron quantum dot", Phys. Rev. Letters, **91**, 196802 (2003).

[22] M. Ikezawa, B. Pal, Y. Masumoto, I. V. Ignatiev, S. Yu Verbin and I. Ya. Gerlovin, "Sub-millisecond electron spin relaxation in InP quantum dots", Phys. Rev. B, **72**, 153302 (2005).

[23] R. Hanson, L. H. Willems vand Beveren, I. T. Vink, J. M. Elzerman, W. J. M. Naber, F. H. L. Koppens, L. P. Kouwenhoven and L. M. K. Vandersypen, "Single-shot readout of electron spin states in a quantum dot using spin dependent tunnel rates", Phys. Rev. Lett., **94**, 196802 (2005).

[24] M. Kroutvar, Y. Ducommun, D. Heiss, M. Bichler, D. Schuh, G. Abstreiter and J. J. Finley, "Optically programmable electron spin memory using semiconductor quantum dots", Nature (London), **432**, 81 (2004).

[25] S. Amasha, K. MacLean, Iuliana Radu, D. M. Zümbuhl, M. A. Kastner, M. P. Hanson and A. C. Gossard, "Measurements of the spin relaxation rate at Low magnetic fields in a quantum dot", www.arXiv.org/cond-mat/0607110.

[26] S. Pramanik, C-G Stefanita, S. Patibandla, S. Bandyopadhyay, K. Garre, N. Harth and M. Cahay, "Observation of extremely long spin

relaxation time in an organic nanowire spin valve", Nature Nanotech., **2**, 216 (2007).

[27] S. Sanvito, "Organic electronics: Memoirs of a spin", Nature Nanotech., **2**, 204 (2007).

[28] S. Cortez, et al., "Optically driven spin memory in n-doped InAs-GaAs quantum dots", Phys. Rev. Letters, **89**, 207401 (2002).

[29] V. N. Golovach, A. Khaetski and D. Loss, "Phonon induced decay of the electron spin in quantum dots", Phys. Rev. Lett., **93**, 016601 (2004).

[30] R. de Sousa and S. Das Sarma, "Electron spin coherence in semiconductors: Considerations for a spin-based solid-state quantum computer architecture", Phys. Rev. B, **67**, 033301 (2003).

[31] S. Bandyopadhyay, et al., "Electrochemically assembled quasi perioidic quantum dot arrays", Nanotechnology, **7**, 360 (1996).

[32] A. Abragam, *The Principles of Nuclear Magnetism*, (Oxford University Press, London, 1961); C. P. Slichter, *Principles of Magnetic Resonance*, (Springer-Verlag, Berlin, 1996).

[33] T. A. Kennedy, J. S. Colton, J. E. Butler, R. C. Linares and P. J. Doering, "Long coherence times at 300 K for nitrogen-vacancy center spins in diamond grown by chemical vapor deposition", Appl. Phys. Lett., **83**, 4190 (2003).

[34] T. Gaebel, et al., "Room-temperature cpherent coupling of single spins in diamond", Nature Physics, **2**, 408 (2006).

[35] M. V. Gurudev Dutt, L. Childress, L. Jiang, E. Togan, J. Maze, F. Jelezko, A. S. Zibrov, P. R. Hemmer and M. D. Lukin, "Quantum register based on individual electronic and nuclear spin qubits in diamond", Science, **316**, 1312 (2007).

[36] J. A. Gupta, D. D. Awschalom, X. Peng and A. P. Alivisatos, "Spin coherence in semiconductor quantum dots", Phys. Rev. B, **59**, R10421 (1999).

[37] R. I. Dzhioev, V. L. Korenev, I. A. Merkulov, B. P. Zakharchenya, D. Gammon, Al. L. Efros and D. S. Katzer, "Manipulation of the spin memory of electrons in n-GaAs", Phys. Rev. Lett., **88**, 256801 (2002).

[38] S. Pramanik, B. Kanchibotla and S. Bandyopadhyay, "Transverse spin relaxation time in an ensemble of electrochemically self assembled CdS quantum dots", Proc. of IEEE NANO 2006, Cincinnati, OH, July 2006 (IEEE Press).

[39] I. A. Merkulov, Al. L. Efros and M. Rosen, "Electron spin relaxation by nuclei in semiconductor quantum dots", Phys. Rev. B, **65**, 205309 (2002).

[40] A. V. Khaetski, D. Loss and L. Glazman, "Electron spin decoherence in quantum dots due to interaction with nuclei", Phys. Rev. Lett., **88**, 186802 (2002).

[41] A. Greilich, A. Shabaev, D. R. Yakovlev, Al. L. Efros, I. A. Yugova, D. Reuter, A. D. Wieck and M. Bayer, "Nuclei induced frequency focusing of electron spin coherence", Science, **317**, 1896 (2007).

[42] A. M. Tyryshkin, S. A. Lyon, W. Jantsch and F. Schaffler, "Spin manipulation of free two dimensional electrons in Si/SiGe quantum wells", Phys. Rev. Lett., **94**, 126802 (2005).

[43] S. Bandyopadhyay and M. Cahay, "Rashba effect in an asymmetric quantum dot in a magnetic field", Superlat. Microstruct., **32**, 171 (2002).

[44] A. S. Bracker, E. A. Stinaff, D. Gammon, M. E. Ware, J. G. Tischler, A. Shabaev, Al. L. Efros, D. Park, D. Gershoni, V. L. Korenev and I. A. Merkulov, "Optical pumping of the electronic and nuclear spin of single charge-tunable quantum dots", Phys. Rev. Lett., **94**, 047402 (2005).

[45] M. V. Gurudev Dutt, Jun Cheng, Bo Li, Xiadong Xu, P. R. Berman, D. G. Steel, A. S. Bracker, D. Gammon, Sophia E. Economou, Ren Bao Liu, and L. J. Sham, "Stimulated and spontaneous optical generation of electron spin coherence in charged GaAs quantum dots", Phys. Rev. Lett., **94**, 227403 (2005).

[46] S. D. Ganichev, E. L. Ivchenko, V. V. Bel'kov, S. A. Tarasenko, M. Sollinger, D. Weiss, W. Wegscheider and W. Prettl, "Spin-galvanic effect", Nature (London), **417**, 153 (2002).

[47] F. Meier and B. P. Zakharchenya, Eds. *Optical Orientation*, (Elsevier Science, Amsterdam, 1984).

9

Exchange Interaction

In this chapter, we discuss the extremely important notion of exchange interaction between electrons, which plays a major role in the operation of many spin-based devices (both classical – see Chapter 13 – and quantum mechanical – see Chapter 14, later). We start by describing the formulation of the *Exclusion Principle* enunciated by Pauli in 1924. We then use this principle to calculate the energy eigenvalues of the Helium atom based on perturbation theory. This leads to the concept of exchange energy which was extended by Heisenberg to a system of spins to propose a model for the origin of ferromagnetism. The concept of exchange coupling and Heisenberg Hamiltonian are crucial in understanding the operation of many solid state spin-based devices described in this book and elsewhere.

9.1 Identical particles and the Pauli Exclusion Principle

In the Hydrogen atom (recall Chapter 1), if we neglect spin-orbit coupling, then the behavior of the single electron is completely described by specifying four quantum numbers (n, l, m, s) that are associated with total energy, orbital angular momentum, z-component of the angular momentum in a magnetic field, and the spin, respectively. When these four quantum numbers are given, the state of the one-electron system is precisely specified. This is one of the few examples in quantum mechanics where the problem is exactly solvable and one can describe the dynamics of the lone electron rigorously.

The quantum mechanics of multi-electron systems do not typically lend themselves to an exact solution. The description of all such systems is imprecise in the sense that the problem cannot be solved exactly. There are however important principles and theorems that govern many electron systems. Probably, the most fundamental of them is the Pauli Exclusion Principle that applies to any two Fermions.

As early as 1924, while analyzing spectroscopic data of atoms with more than one electron, Pauli conjectured that no two electrons can occupy the same quantum state if their wavefunctions overlap. This led him to formulate the famous *Exclusion Principle* that bears his name. This Principle states that *no two Fermions, whose wavefunctions have non-zero overlap, can have*

exactly the same set of quantum numbers. One of the early successes of the Pauli Exclusion Principle was to provide an explanation for the periodic table of the elements.

The concept of exchange energy follows from the Pauli Exclusion Principle. Any determination of the energy eigenvalues and corresponding eigenstates of a system composed of more than one electron must take into account the Coulomb interaction between electrons since electrons are charged particles. Furthermore, for such a system of identical and indistinguishable Fermi particles, the *Symmetry Principle* dictates that the overall wavefunction of the system must be antisymmetric under the operation of swapping the indices of any two electrons. This principle, along with the Pauli Exclusion Principle, gives rise to the exchange interaction.

9.1.1 The Helium atom

The Pauli Exclusion Principle was first applied to study the simplest many-electron system, namely the Helium atom, which has two electrons orbiting a nucleus. Using first-order perturbation theory, one can estimate the energy levels of the Helium atom. Although the use of perturbation theory is crude, it is the simplest way to introduce the concept of exchange energy and illustrate its fundamental connection to "spin" and the Pauli Exclusion Principle. This approach clearly brings out the truly quantum-mechanical origin of the exchange energy concept.

Consider a Helium atom composed of a heavy nucleus with charge $+2e$ and two surrounding electrons, as shown in Figure 9.1. Neglecting the motion of the nucleus and using the latter as the origin of a fixed frame of reference, the Hamiltonian of the Helium atom, taking into account the Coulomb attraction between each electron and the nucleus and the Coulomb repulsion between the electrons, is given by

$$H_{He} = \frac{|\vec{p}_1|^2}{2m_0} + \frac{|\vec{p}_2|^2}{2m_0} - \frac{2e^2}{4\pi\epsilon_0|\vec{r}_1|} - \frac{2e^2}{4\pi\epsilon_0|\vec{r}_2|} + \frac{e^2}{4\pi\epsilon_0|\vec{r}_{12}|}, \tag{9.1}$$

where $\vec{p}_1$ and $\vec{p}_2$ are the momentum operators associated with electrons 1 and 2, respectively. Their spatial coordinates are denoted by $\vec{r}_1$ and $\vec{r}_2$, and $\vec{r}_{12} = \vec{r}_1 - \vec{r}_2$.

Let P be the operator which permutes the two identical electrons. It must commute with the Hamiltonian H_{He} since the latter is invariant upon interchange of the indices 1 and 2; therefore,

$$[H_{He}, P] = 0. \tag{9.2}$$

This implies that the wavefunctions of the Helium atom must be both eigenstates of H_{He} and P. If we call $\phi(e_1, e_2)$ the wavefunction describing the system of two electrons so that $\phi(e_1, e_2)$ satisfies the Schrödinger equation

$$H_{He}\phi(e_1, e_2) = E\phi(e_1, e_2), \tag{9.3}$$

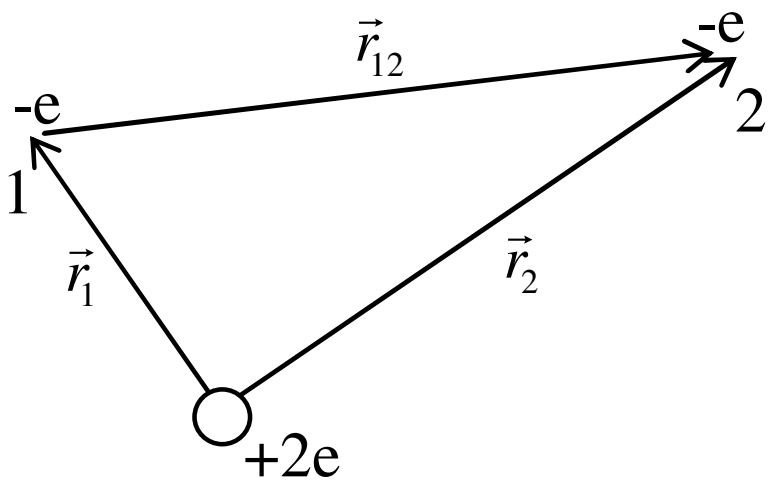


FIGURE 9.1
The Helium atom: The positions of the nucleus and the two electrons are represented.

then we must have

$$P\phi(e_1, e_2) = \phi(e_2, e_1) = \lambda\phi(e_1, e_2), \tag{9.4}$$

where the labels (e_1, e_2) are the set of variables $(\vec{r_1}, s_1)$ and $(\vec{r_2}, s_2)$ representing the position and spin coordinates of electrons 1 and 2, respectively.

Applying the permutation operator to both sides of the last equation, we get

$$P^2\phi(e_1, e_2) = \lambda P\phi(e_1, e_2) = \lambda^2\phi(e_1, e_2) = \phi(e_1, e_2), \tag{9.5}$$

which leads to

$$\lambda^2 = 1, \tag{9.6}$$

or

$$\lambda = \pm 1. \tag{9.7}$$

If $\lambda = +1$, the overall wavefunction $\phi(e_1, e_2)$ is symmetric. On the other hand, if $\lambda = -1$, the wavefunction is antisymmetric. Since we are dealing with a system of Fermions, the Symmetry Principle constrains us to accept the antisymmetric solution only. Furthermore, since we have neglected any spin-orbit interaction, which is the only term which would couple the position and spin components (i.e., the orbital and spin degrees of freedom), the overall wavefunction can be written as a tensor product of a purely spatial part and a

purely spinor part (since the Hamiltonian does not contain any spin-dependent operator). This is the same "variable separation" approach that we visited in Chapter 7. There, it was inexact since we were dealing with spin-orbit interaction. Here, it is permissible since spin-orbit interaction is ignored.

The overall wavefunction $\phi_A(\overrightarrow{r_1}, s_1; \overrightarrow{r_2}, s_2)$ must be antisymmetric upon interchanging the indices 1 and 2, i.e., either

$$\phi_A(\overrightarrow{r_1}, s_1; \overrightarrow{r_2}, s_2) = \psi_S(\overrightarrow{r_1}, \overrightarrow{r_2})\Xi_A(s_1, s_2), \tag{9.8}$$

or

$$\phi_A(\overrightarrow{r_1}, s_1; \overrightarrow{r_2}, s_2) = \psi_A(\overrightarrow{r_1}, \overrightarrow{r_2})\Xi_S(s_1, s_2), \tag{9.9}$$

where the subscripts S and A stand for "symmetric" and "antisymmetric", respectively.

Since each component of the total spin operator

$$\vec{S} = \vec{S_1} + \vec{S_2} = \frac{\hbar}{2}(\vec{\sigma_1} + \vec{\sigma_1}) \tag{9.10}$$

and its square magnitude $S^2 = \vec{S}.\vec{S}$ are conserved quantities (recall Chapter 2), the spinorial part must be eigenstates of these operators. However, since the three individual components of $\vec{S}$ do not commute, but each of them commutes with S^2 (recall Chapter 2 material), we can only select one component of $\vec{S}$ to form a complete set of spinorial operators with S^2. Hereafter, we select the z-component, but the ensuing arguments would hold true if we had selected either S_x or S_y instead of S_z.

As shown in Chapter 2, the eigenvectors of S_z or the Pauli σ_z matrix are given by

$$|0> = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \tag{9.11}$$

with eigenvalue $+1$, and

$$|1> = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \tag{9.12}$$

with eigenvalue -1. We need to add an extra index to these eigenvectors to label which one is associated with which electron in the subsequent steps.

The eigenstates of the spinorial part of the wavefunction for the two electron system are given explicitly in Table 9.1. In this Table, the single wavefunction $\Xi_A(s_1, s_2)$ is referred to as the *singlet* state whereas the three wavefunctions $\Xi_S(s_1, s_2)$ are referred to as the *triplet* states of the Helium atom.

Exercise: Show explicitly that the application of the operators S_z and S^2 to the spinorial states of the two electron system described here are the values listed in columns 2 and 3 in Table 9.1.

TABLE 9.1
Eigenstates of the spinorial part

Spinorial part	S_z	S^2
$\Xi_S(s_1,s_2) = \|0>_1 \|0>_2$	$+\hbar$	$2\hbar^2$
$\Xi_S(s_1,s_2) = \frac{1}{\sqrt{2}}\left[\|0>_1 \|1>_2 + \|0>_2 \|1>_1\right]$	0	$2\hbar^2$
$\Xi_S(s_1,s_2) = \|1>_1 \|1>_2$	$-\hbar$	$2\hbar^2$
$\Xi_A(s_1,s_2) = \frac{1}{\sqrt{2}}\left[\|0>_1 \|1>_2 - \|0>_2 \|1>_1\right]$	0	0

The total spin operator associated with the two-electron system is given by

$$\overrightarrow{S} = \overrightarrow{S_1} + \overrightarrow{S_2} = \frac{\hbar}{2}\overrightarrow{\sigma_1} + \frac{\hbar}{2}\overrightarrow{\sigma_2}, \tag{9.13}$$

with

$$\overrightarrow{\sigma} = (\sigma_x, \sigma_y, \sigma_z), \tag{9.14}$$

and one extra label must be added to distinguish between electron 1 and 2.

Therefore,

$$S^2 = \overrightarrow{S} \cdot \overrightarrow{S} = \frac{\hbar^2}{4}\left(\sigma_1^2 + \sigma_2^2 + 2\sigma_1 \cdot \sigma_2\right) = \frac{\hbar^2}{2}\left(3\times I + \overrightarrow{\sigma_1} \cdot \overrightarrow{\sigma_2}\right), \tag{9.15}$$

where

$$\overrightarrow{\sigma_1} \cdot \overrightarrow{\sigma_2} = \sigma_{1x}\sigma_{2x} + \sigma_{1y}\sigma_{2y} + \sigma_{1z}\sigma_{2z}. \tag{9.16}$$

Since according to the properties of the Pauli spin matrices derived in Chapter 2, we have

$$\begin{aligned} \sigma_x|0\rangle = |1\rangle; \sigma_y|0\rangle = i|1\rangle; \sigma_z|0\rangle = |0\rangle; \\ \sigma_x|1\rangle = |0\rangle; \sigma_y|1\rangle = -i|0\rangle; \sigma_z|1\rangle = -|1\rangle, \end{aligned} \tag{9.17}$$

we get

$$\overrightarrow{\sigma_1} \cdot \overrightarrow{\sigma_2}\{|0\rangle_1|0\rangle_2\} = |0\rangle_1|0\rangle_2. \tag{9.18}$$

Hence

$$\begin{aligned} S^2|0\rangle_1|0\rangle_2 &= \frac{\hbar^2}{2}\left(3\times I + \overrightarrow{\sigma_1} \cdot \overrightarrow{\sigma_2}\right)|0\rangle_1|0\rangle_2 \\ &= 2\hbar^2|0\rangle_1|0\rangle_2. \end{aligned} \tag{9.19}$$

Furthermore,

$$S_z|0\rangle_1|0\rangle_2 = \frac{\hbar}{2}[\sigma_{1z}|0\rangle_1|0\rangle_2 + \sigma_{2z}|0\rangle_1|0\rangle_2] = \hbar|+\rangle_1|+\rangle_2. \tag{9.20}$$

Similar derivations lead to the other eigenvalues listed in Table 9.1.

In reality, the spatial part of the wavefunction must be found through a rigorous solution of the Schrödinger equation

$$H_{He}\left(\overrightarrow{r_1},\overrightarrow{r_2}\right)\psi_{S,A}\left(\overrightarrow{r_1},\overrightarrow{r_2}\right)=E\psi_{S,A}\left(\overrightarrow{r_1},\overrightarrow{r_2},\right). \tag{9.21}$$

Despite the rather simple form of the Hamiltonian, this problem has no known exact solution. Hereafter, we seek solutions to this equation using perturbation theory under the assumption that the Coulomb repulsion between the two electrons is a weak perturbation on the unperturbed Hamiltonian H_0 given by

$$H_0 = H_1 + H_2, \tag{9.22}$$

where

$$H_1 = \frac{|\vec{p}_1|^2}{2m_0} - \frac{2e^2}{4\pi\epsilon_0|\vec{r}_1|}, \tag{9.23}$$

and

$$H_2 = \frac{|\vec{p}_2|^2}{2m_0} - \frac{2e^2}{4\pi\epsilon_0|\vec{r}_2|}. \tag{9.24}$$

Here, H_1 and H_2 are the Hamiltonians of the individual electrons with a kinetic term and a potential energy term due to the attraction of the nucleus with charge $Z = +2e$.

The eigenfunctions and corresponding eigenvalues of H_1 and H_2 can be easily found. They are the well-known wavefunctions for the Hydrogen atom (the Laguerre polynomials for the radial part and the spherical harmonics for the angular parts, if we adopt spherical coordinates) [1] with the only difference being that the charge of the nucleus is $+2e$ rather than $+e$. Once that difference is accounted for, if we label the eigenfunctions and corresponding eigenvalues as $\phi_i(\overrightarrow{r})$ and ϵ_i, respectively, then any product of the form

$$\phi_i(\overrightarrow{r_1})\phi_j(\overrightarrow{r_2}) \tag{9.25}$$

is an eigenfunction of H_0 with eigenvalue $\epsilon_i + \epsilon_j$. A solution of this type is often called the *Hartree* solution after Hartree.

There is one serious problem with the Hartee solution. It does not account for the Symmetry Principle which mandates that the total wavefunction (spatial and spinorial parts together) must be antisymmetric under exchange of indices. In order to reconcile with the Symmetry Principle, we must select a proper linear combination of products of the form (9.25) such that the overall spatial part of the wavefunction of H_0 is either symmetric or antisymmetric. Of course, when the spatial part is symmetric, the spin part must be antisymmetric, and vice versa, so that the overall wavefunction (spatial + spin) is antisymmetric, as mandated by the Symmetry Principle.

If $i = j$, the only possibility is a symmetric spatial wavefunction of the form

$$\psi_S(\overrightarrow{r_1},\overrightarrow{r_2}) = \phi_i(\overrightarrow{r_1})\phi_i(\overrightarrow{r_2}). \tag{9.26}$$

If $i \neq j$, the spatial part of the wavefunction can be either symmetric

$$\psi_S(\overrightarrow{r_1}, \overrightarrow{r_2}) = \frac{1}{\sqrt{2}} \left[\phi_i(\overrightarrow{r_1})\phi_j(\overrightarrow{r_2}) + \phi_i(\overrightarrow{r_2})\phi_j(\overrightarrow{r_1})\right], \tag{9.27}$$

or antisymmetric

$$\psi_A(\overrightarrow{r_1}, \overrightarrow{r_2}) = \frac{1}{\sqrt{2}} \left[\phi_i(\overrightarrow{r_1})\phi_j(\overrightarrow{r_2}) - \phi_i(\overrightarrow{r_2})\phi_j(\overrightarrow{r_1})\right]. \tag{9.28}$$

Excercise: (a) Prove that the factors $1/\sqrt{2}$ appearing in Equations (9.27) and (9.28) are needed to normalize $\psi_S(\overrightarrow{r_1}, \overrightarrow{r_2})$ and $\psi_A(\overrightarrow{r_1}, \overrightarrow{r_2})$. In other words,

$$\int d\overrightarrow{r_1} \int d\overrightarrow{r_2} \psi_s^\star \left(\overrightarrow{r_1}, \overrightarrow{r_2}\right) \psi_s \left(\overrightarrow{r_1}, \overrightarrow{r_2}\right) = 1 \tag{9.29}$$

and a similar relation holds for $\psi_A \left(\overrightarrow{r_1}, \overrightarrow{r_2}\right)$, if the individual ϕ_i's are normalized, i.e.,

$$\int d\overrightarrow{r_1} \phi_i^\star \left(\overrightarrow{r_1}\right) \phi_i \left(\overrightarrow{r_1}\right) = \int d\overrightarrow{r_2} \phi_i^\star \left(\overrightarrow{r_2}\right) \phi_i \left(\overrightarrow{r_2}\right) = 1. \tag{9.30}$$

(b) Show that $\psi_s \left(\overrightarrow{r_1}, \overrightarrow{r_2}\right)$ and $\psi_A \left(\overrightarrow{r_1}, \overrightarrow{r_2}\right)$ are eigenstates of H_0 with energy $\varepsilon_i + \varepsilon_j$. The solution of this exercise is left to the reader.

If the two electrons have the same spatial part, the only possible choice for the spinorial part of the wavefunction which will make the overall wavefunction antisymmetric is the singlet state in Table 9.1, i.e.,

$$\phi_A(\overrightarrow{r_1}, s_1; \overrightarrow{r_2}, s_2) = \frac{1}{\sqrt{2}} \phi_i(\overrightarrow{r_1})\phi_i(\overrightarrow{r_2}) \left[|0>_1 |0>_2 - |0>_2 |0>_1\right]. \tag{9.31}$$

Using first order perturbation theory to find the approximate values of the energies of the Helium atom for which the Coulomb repulsion between electrons is assumed to be small, the energy associated with $\phi_A(\overrightarrow{r_1}, s_1; \overrightarrow{r_2}, s_2)$ is given by

$$E_i = 2\epsilon_i + E_C, \tag{9.32}$$

where the correcting term

$$E_C = \left\langle \phi_A \left| \frac{e^2}{4\pi\epsilon_0 |\vec{r}_{12}|} \right| \phi_A \right\rangle = \left\langle \phi_A \left| \frac{e^2}{4\pi\epsilon_0 |\vec{r}_1 - \vec{r}_2|} \right| \phi_A \right\rangle \tag{9.33}$$

is the Coulomb energy

$$E_C = \frac{e^2}{2} \int d\overrightarrow{r_1} \int d\overrightarrow{r_2} \left[{}_1\langle 0|_2\langle 1| - {}_2\langle 0|_1\langle 1|\right] \frac{|\phi_i(\overrightarrow{r_1})|^2 |\phi_i(\overrightarrow{r_2})|^2}{4\pi\epsilon_0 |\vec{r}_{12}|} \left[|0\rangle_1 |1\rangle_2 - |0\rangle_2 |1\rangle_1\right]. \tag{9.34}$$

As easily checked, the inner product involving the spinorial part of the wavefunction is

$$({}_1\langle 0|_2\langle 1| - {}_2\langle 0|_1\langle 1|)\,(|0\rangle_1|1\rangle_2 - |0\rangle_2|1\rangle_1) = 2. \tag{9.35}$$

Therefore, E_C becomes

$$E_C = \int d\overrightarrow{r_1}\int d\overrightarrow{r_2}\frac{e|\phi_i(\overrightarrow{r_1})|^2|\cdot e\phi_i(\overrightarrow{r_2})|^2}{4\pi\epsilon|\vec{r}_{12}|} = \int d\overrightarrow{r_1}\int d\overrightarrow{r_2}\frac{e|\phi_i(\overrightarrow{r_1})|^2|\cdot e\phi_i(\overrightarrow{r_2})|^2}{4\pi\epsilon|\vec{r}_1 - \vec{r}_2|}, \tag{9.36}$$

which is clearly an energy resulting from the Coulomb repulsion between the charge distributions associated with the probability densities $|\phi_i(\overrightarrow{r_1})|^2$ and $|\phi_i(\overrightarrow{r_2})|^2$.

Exercise: Calculate the value of the Coulomb term E_C if $\phi_i(\overrightarrow{r_1})$, $\phi_i(\overrightarrow{r_2})$ are given by the ground state wavefunctions of the Hydrogen atom. Do not forget to replace the atomic number of the nucleus by $Z = +2e$.

Solution

The integral appearing in Equation (9.36) is more readily performed using the following Fourier transform

$$\frac{1}{r_{12}} = \frac{1}{2\pi^2}\int d\overrightarrow{k}\,\frac{e^{i\,\overrightarrow{k}\cdot(\overrightarrow{r_1}-\overrightarrow{r_2})}}{k^2}. \tag{9.37}$$

The wavefunction for the ground state of the Hamiltonians H_1 and H_2 in Equations (9.23) and (9.24) is

$$\phi_{1s}(r) = \frac{1}{\sqrt{\pi a_{He}^3}}e^{-\frac{r}{a_{He}}}, \tag{9.38}$$

where $a_{He} = a_0/2$ and a_0 is the Bohr radius.

Hence

$$E_C = \frac{1}{4\pi\varepsilon_0}\frac{e^2}{2\pi^2}\int\frac{d\,\overrightarrow{k}}{k^2}\left(\int d\overrightarrow{r_1}e^{i\,\overrightarrow{k}\cdot\overrightarrow{r_1}}|\phi_{1s}(r_1)|^2\right)\left(\int d\overrightarrow{r_2}e^{-i\,\overrightarrow{k}\cdot\overrightarrow{r_2}}|\phi_{1s}(r_2)|^2\right), \tag{9.39}$$

or

$$E_C = \frac{1}{4\pi\varepsilon_0}\frac{e^2}{2\pi^2}\int\frac{d\,\overrightarrow{k}}{k^2}g\left(\overrightarrow{k}\right)g^\star\left(\overrightarrow{k}\right), \tag{9.40}$$

where the asterisk represents complex conjugate and $g(\overrightarrow{k})$ is given by

$$g\left(\overrightarrow{k}\right) = \int d\overrightarrow{r}e^{i\,\overrightarrow{k}\cdot\overrightarrow{r}}|\phi_{1s}(r)|^2. \tag{9.41}$$

Since $|\phi_{1s}(r)|^2$ depends only on the magnitude of r, the vector can be selected to be along the z-axis and we get

$$g\left(\overrightarrow{k}\right) = \int_0^{+\infty} dr r^2 2\pi \int_0^{\pi} d\theta \sin\theta e^{ikr\cos\theta} |\phi_{1s}(r)|^2. \tag{9.42}$$

The latter can be easily calculated using Equation (9.38) and is found to be

$$g\left(\overrightarrow{k}\right) = \frac{16}{\left(4 + a_{He}^2 k^2\right)^2}. \tag{9.43}$$

Therefore, the Coulomb energy is given by

$$\begin{aligned} E_C &= \frac{1}{4\pi\varepsilon_0}\frac{e^2}{2\pi^2}\int \frac{d\overrightarrow{k}}{k^2}\left(\frac{16}{\left(4 + a_{He}^2 k^2\right)^2}\right)^2 \\ &= \frac{5e^2}{32\pi\varepsilon_0 a_0}. \end{aligned} \tag{9.44}$$

Since the ground state of the Hamiltonian H_0 in Equation (9.22) has an energy equal to $+8E_{1s}$, where E_{1s} is the energy of the ground state of the Hydrogen atom

$$E_{1s} = -\frac{1}{4\pi\varepsilon_0}\frac{e^2}{2a_0}, \tag{9.45}$$

we have

$$\frac{E_C}{8E_{1s}} = \frac{-5}{16} \approx -0.3. \tag{9.46}$$

Despite the fact that E_C is not quite a small correction to the unperturbed value $8E_{1s}$, the derivation given above is quite well suited to describe the concept of exchange energy when considering the case where the spatial part of the electron wavefunctions are different, as described next.

If the electron wavefunctions $\phi_i(\overrightarrow{r})$, $\phi_j(\overrightarrow{r})$ are different, i.e., $i \neq j$, then the overall wavefunction of the total Hamiltonian H_{He} can be of two forms. If we choose the spinorial part of the overall wavefunction to be one of the triplet state, the spatial part must be antisymmetric to keep the overall wavefunction antisymmetric. The simplest way to write the spatial part is

$$\psi_{2-electron} = \frac{1}{\sqrt{2}}\left[\phi_i(\overrightarrow{r_1})\phi_j(\overrightarrow{r_2}) - \phi_i(\overrightarrow{r_2})\phi_j(\overrightarrow{r_1})\right]. \tag{9.47}$$

This type of wavefunction is often referred to as *Slater determinant* after Slater:

$$\psi_{2-electron} = \frac{1}{\sqrt{2}} det \begin{vmatrix} \phi_i(\overrightarrow{r_1}) & \phi_i(\overrightarrow{r_2}) \\ \phi_j(\overrightarrow{r_1}) & \phi_j(\overrightarrow{r_2}) \end{vmatrix}. \tag{9.48}$$

In the theory of the Helium atom, the triplet states composed of the product of this spatial part and the three spin states described as triplet states in Table 9.1 are referred to as the *orthohelium* states.

If, on the other hand, we select the spinorial part of the overall wavefunction to be the singlet state in Table 9.1, the spatial part must be symmetric and the simplest way to achieve this is to write the spatial part as follows

$$\frac{1}{\sqrt{2}}\left[\phi_i(\overrightarrow{r_1})\phi_j(\overrightarrow{r_2}) + \phi_i(\overrightarrow{r_2})\phi_j(\overrightarrow{r_1})\right]. \tag{9.49}$$

In the theory of the Helium atom, the states written as the product of symmetrical spatial part and antisymmetric spinorial part are referred to as *parahelium* states.

When $i \neq j$, the energy associated with the Hamiltonian H_0 are given by $\epsilon_i + \epsilon_j$. For the overall Hamiltonian containing the effects of the Coulomb repulsion, we have, to the first order in perturbation theory, a correction given by

$$E_S = \frac{1}{2}\int d\overrightarrow{r_1}\int d\overrightarrow{r_2}\frac{\left(\phi_i^*(\overrightarrow{r_1})\phi_j^*(\overrightarrow{r_2}) + \phi_j^*(\overrightarrow{r_1})\phi_i^*(\overrightarrow{r_2})\right)\left(\phi_i(\overrightarrow{r_1})\phi_j(\overrightarrow{r_2}) + \phi_j(\overrightarrow{r_1})\phi_i(\overrightarrow{r_2})\right)}{4\pi\epsilon_0|\vec{r}_{12}|}$$
$$\times\langle\Xi_A(\vec{r}_1,\vec{r}_2)|\Xi_A(r_1,r_2)\rangle \tag{9.50}$$

when the two electrons are in the singlet state.

Using $\Xi_A(\lambda_1,\lambda_2)$ listed in Table 9.1, the vector product is easily shown to be equal to unity leading to

$$E_S = K_{ij} + J_{ij}, \tag{9.51}$$

where

$$K_{ij} = \int d\overrightarrow{r_1}\int d\overrightarrow{r_2}\frac{e|\phi_i(\overrightarrow{r_1})|^2\cdot e\phi_j(\overrightarrow{r_2})|^2}{4\pi\epsilon_0|\vec{r}_{12}|} \tag{9.52}$$

is the Coulomb energy term found earlier for the case $i = j$, and

$$J_{ij} = e^2\int d\overrightarrow{r_1}\int d\overrightarrow{r_2}\frac{\phi_i^*(\overrightarrow{r_1})\phi_j^*(\overrightarrow{r_2})\phi_j(\overrightarrow{r_1})\phi_i(\overrightarrow{r_2})}{4\pi\epsilon_0|\vec{r}_{12}|}. \tag{9.53}$$

This last term is referred to as the *exchange energy* term and is truly quantum-mechanical in origin. Its appearance can be traced back to the Pauli exclusion principle and to the fact that the two electrons in the Helium atom are indistinguishable and therefore must obey the Symmetry Principle. For two such particles in close proximity, their wavefunction will overlap and if we introduce an overlap density for the system of these two electrons as follows

$$\rho_{ij}(\overrightarrow{r_1}) = \phi_i^*(\overrightarrow{r_1})\phi_j(\overrightarrow{r_1}), \tag{9.54}$$

and

$$\rho_{ij}(\overrightarrow{r_2}) = \phi_j^*(\overrightarrow{r_2})\phi_i(\overrightarrow{r_2}), \tag{9.55}$$

then J_{ij} can be written as

$$J_{ij} = e^2 \int d\overrightarrow{r_1} \int d\overrightarrow{r_2} \frac{\rho_{ij}(\overrightarrow{r_1})\rho_{ij}(\overrightarrow{r_2})}{|\vec{r}_{12}|}, \tag{9.56}$$

which appears in a form similar to the Coulomb term K_{ij} where the true charge density of individual charges is replaced by the overlap charge densities ρ_{ij}.

A similar calculation starting with the overall wavefunction associated with the triplet states leads to

$$E_T = K_{ij} - J_{ij}. \tag{9.57}$$

For the case of the Helium atom, the exchange energy term J_{ij} is positive and the triplet states (orthohelium) possess *lower* energy than the singlet states (parahelium).

Exercise: For the case where the wavefunction of the two electrons is written as a product of Equation (9.47) and $\Xi_S(s_1, s_2)$ in Table 9.1, show that the average value of the Coulomb repulsion term is indeed given by Equation (9.52).

9.1.2 The Heitler-London model of the Hydrogen molecule

In the early days of quantum mechanics, Heitler and London [2] used an approach similar to the one described for the He atom to calculate the energy eigenvalues of the H_2 molecule. Note that the Hydrogen molecule is different from the Helium atom in that here there are two nuclei instead of one. There is one other very significant difference. Contrary to the case of the Helium atom, the lowest energy state is the singlet state, referred to as *parahydrogen*, rather than the triplet states. This is an intriguing difference between the He atom and the H_2 molecule, both of which have two electrons. It is believed that if two electrons are bound to the *same* potential (the lone nucleus in the case of the Helium atom), then the triplet state has lower energy than the singlet state. If, on the other hand, the two electrons are bound to two *different* potentials (two different nuclei in the case of the Hydrogen molecule), then the singlet state has the lower energy.

A solid state analog of the H_2 molecule is a system of two electrons confined to two coupled quantum dots. The confining potential of a quantum dot acts like the confining potential of a proton. Loss and DiVincenzo [14] have used the Heitler-London approach to determine the energy eigenstates and eigenfunctions of two electrons bound to two coupled quantum dots (separated by a tunnel barrier). These dots are delineated electrostatically in a two-dimensional electron gas. Since the exchange energy depends on the overlap between the wavefunctions of the two electrons, it can be tuned by adjusting the width and height of the tunnel barrier between the quantum

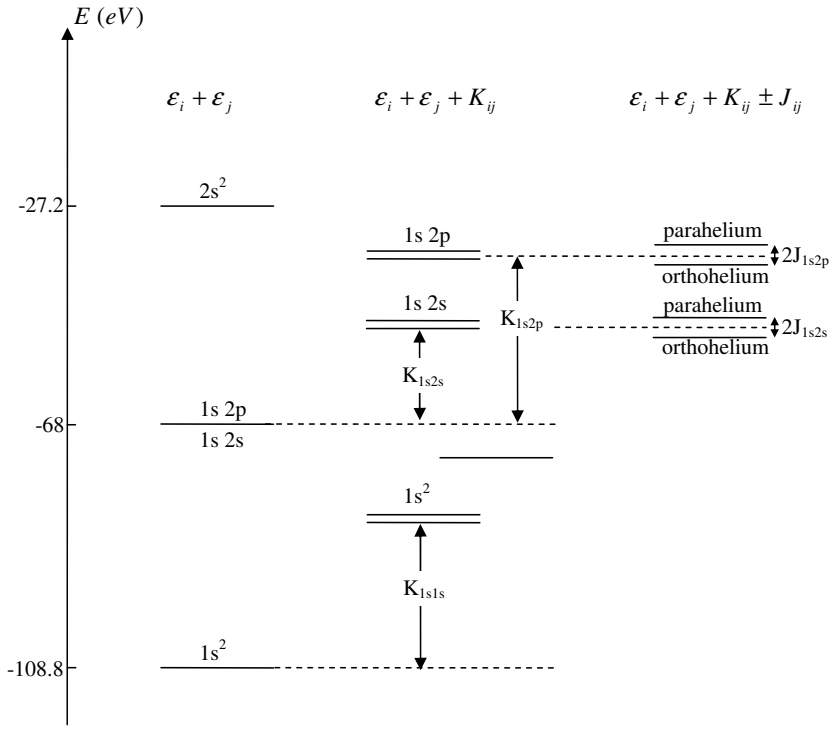


FIGURE 9.2
Schematic representation of the energy level of the Helium atom. From left to right, the energy levels neglecting the effects of the Coulomb interaction between electrons, with the effects of the Coulomb repulsion, and including both the effects of the Coulomb repulsion and the exchange interaction.

dots using metallic Schottky gates. Modulation of the exchange energy in this fashion forms the basis of quantum logic gates (see Chapter 14), particularly the two qubit square-root-of-swap operation. The validity and limitation of the Heitler-London approach to estimate the exchange coupling in semiconductor nanostructures has been assessed recently by several authors [4]. More rigorous numerical calculations of the exchange energy in coupled quantum dots have been presented by Melnikov and Leburton [5]. They solve the two-electron Schrödinger equation self consistently with the Poisson equation in the presence of a magnetic field:

$$\begin{aligned} H\psi\left(\vec{r}_1,\vec{r}_2\right) &= \left[\sum_{i=1}^{2} h_i + \frac{e^2}{4\pi\epsilon\left|\vec{r}_1-\vec{r}_2\right|}\right]\psi\left(\vec{r}_1,\vec{r}_2\right) \\ h_i &= \left(\vec{p}_i - e\vec{A}_i\right)^2/2m^* + V\left(\vec{r}_i\right) \\ \vec{\nabla}\cdot\left(\epsilon\vec{\nabla}V(\vec{r}\right) &= \rho(|\psi|^2), \end{aligned} \tag{9.58}$$

where ϵ is the dielectric constant of the medium, $\vec{A}$ is the vector potential due to any external magnetic field, and ρ is the charge density which depends on the squared magnitude of the wavefunction. The above equations are solved numerically in three dimensions. Melkinov et al. [5] found that the exchange energy is very sensitive to the confining potentials of the quantum dots and barely exceeds 1 meV in realistic quantum dots defined by gates on a two-dimensional electron gas. This does not bode well for spin devices that depend on the modulation of a large exchange energy for their operation.

Since (in units of $\hbar$)

$$\begin{aligned} \overrightarrow{S}_1\cdot\overrightarrow{S}_2 &= \frac{1}{2}\left[\left(\overrightarrow{S}_1+\overrightarrow{S}_2\right)^2 - S_1^2 - S_1^2\right] \\ &= \frac{1}{2}\left[S(S+1) - s_1(s_1+1) - s_2(s_2+1)\right], \end{aligned} \tag{9.59}$$

where $s_1 = s_2 = \frac{1}{2}$.

For triplet states,

$$\left(\overrightarrow{S}_1\cdot\overrightarrow{S}_2\right)_T = \frac{1}{4}, \tag{9.60}$$

and for singlet states

$$\left(\overrightarrow{S}_1\cdot\overrightarrow{S}_2\right)_S = -\frac{3}{4}. \tag{9.61}$$

Hence

$$E_T - E_S = -2J_{12}\left[\frac{1}{4} - \left(\frac{-3}{4}\right)\right] = -2J_{12}. \tag{9.62}$$

Neglecting the energy corresponding to the unperturbed portion of the Hamiltonian H_0 (column 1 in Table 9.1), the correction to first order perturbation theory can be written as follows

$$\Delta_{ij} = K_{ij} - 2\left(\overrightarrow{S}_1 \cdot \overrightarrow{S}_2 + \frac{1}{4}\right) J_{ij}, \tag{9.63}$$

or

$$\Delta_{ij} = \left(K_{ij} - \frac{1}{2} J_{ij}\right) - 2\overrightarrow{S}_1 \cdot \overrightarrow{S}_2 J_{ij}. \tag{9.64}$$

9.2 Hartree and Hartree-Fock approximations

So far, we have mostly dealt with two electrons. But now, let us take a look at true many-body systems comprising numerous electrons. The many-body wavefunction of N interacting electrons is written as

$$\Psi = \Psi\left(\vec{r}_1 s_1, \vec{r}_2 s_2, ...\vec{r}_N s_N\right). \tag{9.65}$$

The *Variational Principle* states that if Ψ is a solution of the many body Schrödinger equation, then the quantity

$$\langle H \rangle = \frac{(\Psi, H\Psi)}{(\Psi, \Psi)} \tag{9.66}$$

must be stationary. Here (A, B) stands for the integral $\int d\vec{r} A^* B$. In fact, the ground state wavefunction is the Ψ which minimizes $\langle H \rangle$.

The Hartree wavefunction will be a product wavefunction:

$$\Psi = \Psi\left(\vec{r}_1 s_1, \vec{r}_2 s_2, ...\vec{r}_N s_N\right) = \psi_1\left(\vec{r}_1 s_1\right) \psi_2\left(\vec{r}_2 s_2\right) ...\psi_N\left(\vec{r}_N s_N\right). \tag{9.67}$$

This wavefunction however does *not* satisfy the Symmetry Principle which mandates that the sign of Ψ should change if we interchange any two of its arguments, i.e.,

$$\Psi\left(\vec{r}_1 s_1, ...\vec{r}_i s_i, ...\vec{r}_j s_j, ...\vec{r}_N s_N\right) = -\Psi\left(\vec{r}_1 s_1, ...\vec{r}_j s_j, ...\vec{r}_i s_i, ...\vec{r}_N s_N\right). \tag{9.68}$$

However, a wavefunction that does satisfy the Symmetry Principle is the *Slater determinant* that we encountered earlier:

$$\Psi\left(\vec{r}_1 s_1, \vec{r}_2 s_2, ...\vec{r}_N s_N\right) = det \begin{pmatrix} \psi_1\left(\vec{r}_1 s_1\right) & \psi_1\left(\vec{r}_2 s_2\right) & \dots & \psi_1\left(\vec{r}_N s_N\right) \\ \psi_2\left(\vec{r}_1 s_1\right) & \psi_2\left(\vec{r}_2 s_2\right) & \dots & \psi_2\left(\vec{r}_N s_N\right) \\ . & . & \dots & . \\ . & . & \dots & . \\ \psi_N\left(\vec{r}_1 s_1\right) & \psi_N\left(\vec{r}_2 s_2\right) & \dots & \psi_N\left(\vec{r}_N s_N\right) \end{pmatrix}. \tag{9.69}$$

This wavefunction automatically satisfies the Pauli principle since it vanishes if $\psi_i = \psi_j$, showing that no two electrons can be put in the same state.

It can be shown that if the energy in Equation (9.66) is evaluated with Ψ given by Equation (9.69) where the ψ-s are orthonormal, then the result is

$$\begin{aligned}\langle H\rangle = &\sum_i \int d\vec{r}\psi_i^* \vec{r}\left(-\frac{\hbar^2}{2m^*}\right)\psi_i\vec{r} \\ &\frac{1}{2}\sum_{i,j}\int d\vec{r}d\vec{r}'\frac{e^2}{4\pi\epsilon\left|\vec{r}-\vec{r}'\right|}\left|\psi_i(\vec{r})\right|^2\left|\psi_j(\vec{r}')\right|^2 \\ &-\frac{1}{2}\sum_{i,j}\int d\vec{r}d\vec{r}'\frac{e^2}{4\pi\epsilon\left|\vec{r}-\vec{r}'\right|}\delta_{s_i s_j}\psi_i^*(\vec{r})\psi_i(\vec{r}')\psi_j^*(\vec{r}')\psi_j(\vec{r}), \quad (9.70)\end{aligned}$$

where the δ is a Krönicker delta indicating that the last term, which is the exchange term, is non-zero only between parallel spins. The second term is the Coulomb term. The exchange term is *negative* and hence tends to lower the total energy. Evaluating $< H >$ with the Slater determinant is known as the *Hartee Fock* approximation. Unlike the bare *Hartree* approximation, it accounts for the Symmetry Principle.

Minimizing the above equation with respect to ψ_i^* leads to the *Hartree Fock Equation* for ψ_i:

$$\begin{aligned}&-\frac{\hbar^2}{2m^*}\psi_i(\vec{r}) + \left[\int d\vec{r}'\left|\psi_i\vec{r}'\right|^2\frac{e^2}{4\pi\epsilon\left|\vec{r}-\vec{r}'\right|}\right]\psi_i(\vec{r}) \\ &-\sum_j\int d\vec{r}'\frac{e^2}{4\pi\epsilon\left|\vec{r}-\vec{r}'\right|}\psi_j^*(\vec{r}')\psi_i(\vec{r}')\psi_j(\vec{r})\delta_{s_i s_j} = E_i\psi_i(\vec{r}). \quad (9.71)\end{aligned}$$

If the ψ-s are plane waves, then the right hand side of the above equation assumes the form $E(\vec{k}_i)\psi_i$, where

$$\begin{aligned}E(\vec{k}) &= \frac{\hbar^2k^2}{2m^*} - \frac{1}{V}\sum_{k'\le k_F}\frac{e^2}{\epsilon\left|\vec{k}-\vec{k}'\right|^2} \\ &= \frac{\hbar^2k^2}{2m^*} - \int_0^{k_F}\frac{d\vec{k}'}{(2\pi)^3}\frac{e^2}{\epsilon\left|\vec{k}-\vec{k}'\right|^2} \\ &= \frac{\hbar^2k^2}{2m^*} - \frac{e^2}{2\pi^2\epsilon}k_F f\left(\frac{k}{k_F}\right), \quad (9.72)\end{aligned}$$

where V is the normalizing volume, k_F is the Fermi wavevector and

$$f(x) = \frac{1}{2} + \frac{1-x^2}{4x}ln\left|\frac{1+x}{1-x}\right|. \quad (9.73)$$

The total energy of N interacting electrons is obtained by summing the energy given by Equation (9.72) over all wavevector states $\vec{k}$ up to the Fermi

wavevector k_F, multiplying by 2 to account for the spin degeneracy, and then dividing the last term by 2 to correct for double counting. This yields

$$E_N = 2\left\{\sum_{k<k_F}\frac{\hbar^2k^2}{2m^*} - \frac{e^2k_F}{4\pi^2\epsilon}\sum_{k<k_F}\left[1+\frac{k_F^2-k^2}{2kk_F}ln\left|\frac{k_F+k}{k_F-k}\right|\right]\right\}. \tag{9.74}$$

The first term in the equation above can be found by converting the summation to an integral:

$$2\sum_{k<k_F}\frac{\hbar^2k^2}{2m^*} = 2\frac{V}{8\pi^3}\int_0^{k_F}\frac{\hbar^2}{2m^*}4\pi k^2dk = \frac{V}{5\pi^2}\frac{\hbar^2}{2m^*}k_F^5. \tag{9.75}$$

Further, noting that the electron density $N/V = k_F^3/3\pi^2$ (this is left as an exercise for the reader), we obtain that

$$2\sum_{k<k_F}\frac{\hbar^2k^2}{2m^*} = \frac{3}{5}\frac{\hbar^2k_F^2}{2m^*}N = \frac{3}{5}NE_F, \tag{9.76}$$

where E_F is the Fermi energy.

The second term in Equation (9.74) can be evaluated by converting the summation to an integral. This exercise is also left to the reader. The final result is

$$E_N = N\left[\frac{3}{5}E_F - \frac{3e^2k_F}{16\pi^2\epsilon}\right]. \tag{9.77}$$

9.3 The role of exchange in ferromagnetism

9.3.1 The Bloch model of ferromagnetism

Felix Bloch [6] pointed out that the last equation, the result of Hartree-Fock approximation, can explain ferromagnetism in a gas of electrons interacting through mutual Coulomb repulsion. In deriving Equation (9.77), we assumed that every state (labeled by the wavevector k) is filled by two electrons of opposite spins, in keeping with the Pauli principle. A more general approach (also consistent with the Pauli principle) is to fill each one-electron level $k < k_\uparrow$ with up-spin electrons and $k < k_\downarrow$ with spin down electrons. Then,

$$\begin{aligned} E_\uparrow &= N_\uparrow\left[\frac{3}{5}\frac{\hbar^2k_\uparrow^2}{2m^*} - \frac{3e^2k_\uparrow}{16\pi^2\epsilon}\right] \\ E_\downarrow &= N_\downarrow\left[\frac{3}{5}\frac{\hbar^2k_\downarrow^2}{2m^*} - \frac{3e^2k_\downarrow}{16\pi^2\epsilon}\right], \end{aligned} \tag{9.78}$$

where

$$E_N = E_\uparrow + E_\downarrow$$
$$\frac{N}{V} = \frac{N_\uparrow}{V} + \frac{N_\downarrow}{V} = \frac{k_\uparrow^3}{6\pi^2} + \frac{k_\uparrow^3}{6\pi^2} = \frac{k_F^3}{3\pi^2}. \tag{9.79}$$

Equation (9.77) is the form E_N takes if $N_\uparrow = N_\downarrow = N/2$. But now let us assume that the electron gas is completely spin polarized so that $N_\uparrow = N$ and $N_\downarrow = 0$. Then $E_N = E_\uparrow$ and $k_\uparrow = 2^{1/3}k_F$. Therefore, in a completely spin polarized gas,

$$E_N^{polarized} = N\left[\frac{3}{5}2^{2/3}\frac{\hbar^2 k_F^2}{2m^*} - 2^{1/3}\frac{3e^2 k_F}{16\pi^2\epsilon}\right]. \tag{9.80}$$

Comparing the above equation with Equation (9.77), we obtain

$$E_N^{polarized} - E_N^{unpolarized} = N\left[\frac{3}{5}\left(2^{2/3}-1\right)\frac{\hbar^2 k_F^2}{2m^*} - \left(2^{1/3}-1\right)\frac{3e^2 k_F}{16\pi^2\epsilon}\right]. \tag{9.81}$$

If this quantity is negative, i.e., if the exchange term (second term) is larger than the first term, then the polarized state, which is *ferromagnetic*, is preferred as the ground state of the electron gas. Generally speaking, low densities (smaller k_F) will favor the ferromagnetic state. Thus, Bloch was able to explain ferromagnetism in an interacting electron gas where electrons interact only via Coulomb repulsion.

9.3.2 The Heisenberg model of ferromagnetism

Independent of Bloch, Heisenberg had postulated another scenario where ferromagnetism could be traced to exchange interaction between electrons. He was the first to propose that the internal magnetic field in ferromagnets, arising from all spins within a domain lining up parallel to each other, could be linked to the presence of exchange coupling between neighboring ionic spins. The basic argument is as follows: From the discussion of the energy eigenstates of the Helium atoms, the difference between energy states (triplet and singlet) associated with orbital (i, j) characterizing the spatial parts of the orbitals is given by

$$E_T - E_S = -2J_{ij}. \tag{9.82}$$

The latter can be formally obtained as eigenvalues of a Hamiltonian

$$H_{ij} = -2J_{ij}\,\overrightarrow{S}_1 \cdot \overrightarrow{S}_2, \tag{9.83}$$

where $\overrightarrow{S}_1$ and $\overrightarrow{S}_2$ are the spin operators (in units of $\hbar$) associated with electrons 1 and 2, i.e., H_{ij} can also be written as follows

$$H_{ij} = -\frac{J_{ij}}{2}\overrightarrow{\sigma}_1 \cdot \overrightarrow{\sigma}_2. \tag{9.84}$$

Since when J_{ij} is positive, a parallel spin configuration (triplet state) has lower energy than the anti-parallel configuration (singlet state), Heisenberg postulated that a generalization of the exchange concept is well suited to describe the onset of ferromagnetism in a system of ionic spins on a lattice. In Heisenberg's model, an ensemble of ionic spins is characterized by a Hamiltonian of the form

$$\widehat{H} = -\frac{1}{2}\sum_{i<j} J_{ij}\,\overrightarrow{\sigma}_i \cdot \overrightarrow{\sigma}_j, \tag{9.85}$$

where the sum is carried out over both indices with $i < j$ to avoid double counting and self-interacting terms.

For given $J'_{ij}s$, if the lowest state of the Hamiltonian corresponds to all spins aligned, then this will account for the onset of ferromagnetism, since the lowest state corresponds to the ground state.

The Heisenberg Hamiltonian is a fundamental concept in the theory of magnetism. Early solutions for the case of interacting spins with only one spin component, i.e., $\overrightarrow{\sigma}_1 \cdot \overrightarrow{\sigma}_2$ replaced by $\sigma_{1z}\sigma_{2z}$ and nearest neighbor interaction ($J_{ij} = J$ for $|i-j| = 1$ and 0 otherwise), were worked out by Ising and Onsager for one and two-dimensional lattices of interacting spins. A complete solution to the full three-dimensional Heisenberg Hamiltonian with the spin operators is still lacking.

9.4 The Heisenberg Hamiltonian

The Heisenberg Hamiltonian in Equation (9.85) is an important theoretical tool in many areas of spintronics. We will have occasion to use it in the description of Single Spin Logic gates in Chapter 13. It is also widely used in the theoretical description of many other spin-based devices that rely on the manipulation of exchange interaction for their operation. One example of this is 2-qubit quantum logic gates discussed in Chapter 14.

This Chapter has introduced the notion of exchange interaction and showed how this notion is at the heart of fundamental physical phenomena such as ferromagnetism. In later Chapters, we will show how this is also at the heart of single spin devices that perform useful operations such as classical or quantum logic.

9.5 Problems

- **Problem 9.1**

  The following problem should be solved using material from Section 2.7.1 in Chapter 2.

  Show that when there are two electrons (or any two spin-1/2 particles), the 2-body wavefunction for which the sum of the two spins equals 1 does not change its value when the spin variables of the electrons are exchanged (that is, the function is symmetric). The wavefunction for which the sum becomes 0 changes sign when the spin variables are interchanged (i.e., the function is antisymmetric).

  The proof of this theorem is taken from S. I. Tomonoga, The Story of Spin (The University of Chicago Press, Chicago, 1997).

  Let the spin operators of the two electrons be $\mathbf{s}_1 = (s_{1x}, s_{1y}, s_{1z})$ and $\mathbf{s}_2 = (s_{2x}, s_{2y}, s_{2z})$. The squared magnitude of the sum is

  $$|\mathbf{s}_1 + \mathbf{s}_2|^2 = |\mathbf{s}_1|^2 + |\mathbf{s}_2|^2 + 2(\mathbf{s}_1 \cdot \mathbf{s}_2), = \frac{1}{2}\left[3 + 4(\mathbf{s}_1 \cdot \mathbf{s}_2)\right], \qquad (9.86)$$

  where we have used Equation (2.17) to derive the last equality.

  The two body wavefunction $\psi(s_{z1}, s_{z2})$ can be written as a column vector spanning all possible spin configurations

  $$\psi_{two-body} = \begin{bmatrix} \psi(+1/2, +1/2) \\ \psi(+1/2, -1/2) \\ \psi(-1/2, +1/2) \\ \psi(-1/2, -1/2) \end{bmatrix}. \qquad (9.87)$$

  In the above equation, the second coordinate of ψ refers to the spin of the second electron and the first coordinate refers to the spin of the first*. We have dropped the coordinate $\mathbf{x}$ since it is not important here.

  We need to calculate $[\mathbf{s_1} \cdot \mathbf{s_2}]\psi$. Note that $\mathbf{s_1} \cdot \mathbf{s_2} = s_{1x}s_{2x} + s_{1y}s_{2y} + s_{1z}s_{2z}$, so that $\mathbf{s_1} \cdot \mathbf{s_2}\psi = [s_{1x}s_{2x} + s_{1y}s_{2y} + s_{1z}s_{2z}]\psi$.

  We now apply the "rules" of Equation (2.70):

*It may appear to the student that when we label one electron as the "first" and the other as the "second" electron, we are somehow distinguishing between indistinguishable particles, which is not allowed. This is actually not the case. We are using these labels only as a matter of convenience. The two electrons are still treated as equals and are indistinguishable.

$$s_{2z}\psi_{two-body} = \begin{bmatrix} (1/2)\psi(+1/2,+1/2) \\ -(1/2)\psi(+1/2,-1/2) \\ (1/2)\psi(-1/2,+1/2) \\ -(1/2)\psi(-1/2,-1/2) \end{bmatrix}, \tag{9.88}$$

$$s_{1z}s_{2z}\psi_{two-body} = \begin{bmatrix} (1/4)\psi(+1/2,+1/2) \\ -(1/4)\psi(+1/2,-1/2) \\ -(1/4)\psi(-1/2,+1/2) \\ (1/4)\psi(-1/2,-1/2) \end{bmatrix}. \tag{9.89}$$

Similarly,

$$s_{1y}s_{2y}\psi_{two-body} = \begin{bmatrix} -(1/4)\psi(-1/2,-1/2) \\ (1/4)\psi(-1/2,+1/2) \\ (1/4)\psi(+1/2,-1/2) \\ -(1/4)\psi(1/2,1/2) \end{bmatrix} \tag{9.90}$$

and

$$s_{1x}s_x\psi_{two-body} = \begin{bmatrix} (1/4)\psi(-1/2,-1/2) \\ (1/4)\psi(-1/2,+1/2) \\ (1/4)\psi(+1/2,-1/2) \\ (1/4)\psi(1/2,1/2) \end{bmatrix}. \tag{9.91}$$

Therefore

$$[\mathbf{s_1}\cdot\mathbf{s_2}]\psi_{two-body} = \begin{bmatrix} (1/4)\psi(1/2,1/2) \\ (1/2)\psi(-1/2,+1/2) - (1/4)\psi(+1/2,-1/2) \\ (1/2)\psi(+1/2,-1/2) - (1/4)\psi(-1/2,+1/2) \\ (1/4)\psi(-1/2,-1/2) \end{bmatrix}. \tag{9.92}$$

If we use Equation (9.86) here, then we get

$$|\mathbf{s_1}+\mathbf{s_2}|^2\psi_{two-body} = \begin{bmatrix} 2\psi(1/2,1/2) \\ \psi(-1/2,+1/2)+\psi(+1/2,-1/2) \\ \psi(+1/2,-1/2)+\psi(-1/2,+1/2) \\ 2\psi(-1/2,-1/2) \end{bmatrix}. \tag{9.93}$$

Now, the total spin = 1 implies that $|\mathbf{s_1}+\mathbf{s_2}|^2\psi_{two-body} = 1(1+1)$ $\psi_{two-body} = 2\psi_{two-body}$, and the total spin = 0 implies $|\mathbf{s_1}+\mathbf{s_2}|^2\psi_{two-body}$

$= 0$. In the former case, if we compare Equations (9.87) and (9.93), then we get

$$\psi(+1/2,-1/2)+\psi(-1/2,+1/2)=2\psi(+1/2,-1/2)$$
$$or,\quad \psi(-1/2,+1/2)=\psi(+1/2,-1/2). \tag{9.94}$$

Therefore,

$$\psi(s_{1z},s_{2z})=\psi(s_{2z},s_{1z}). \tag{9.95}$$

In the latter case, we need

$$\psi(1/2,1/2)=\psi(-1/2,-1/2)=0$$
$$\psi(+1/2,-1/2)+\psi(-1/2,+1/2)=0. \tag{9.96}$$

.

The last equality gives

$$\psi(s_{1z},s_{2z})=-\psi(s_{2z},s_{1z}). \tag{9.97}$$

Equations (9.95) and (9.97) prove the theorem.

- **Problem 9.2**

  Using the theory of the Helium atom in Section 9.1.1, calculate the Coulomb (K_{1s2s}) and exchange (J_{1s2s}) integrals. The ground state wavefunction is given by Equation (9.38) and the first excited state wavefunction is

$$\phi_{2s}(r)=\frac{1}{\sqrt{8\pi a_{He}^3}}\left(1-\frac{r}{2a_{He}}\right)e^{-\frac{r}{2a_{He}}}, \tag{9.98}$$

  where $a_{He}=a_0/2$ and a_0 is the Bohr radius (=0.529 Å).

  Solution

$$K_{1s2s}=\frac{1}{4\pi\varepsilon_0}\frac{e^2}{2\pi^2}\int\frac{d\vec{k}}{k^2}\left(\int d\vec{r}_1 e^{i\vec{k}\cdot\vec{r}_1}|\phi_{1s}(r_1)|^2\right)$$
$$\times\left(\int d\vec{r}_2 e^{-i\vec{k}\cdot\vec{r}_2}|\phi_{2s}(r_2)|^2\right), \tag{9.99}$$

  which yields

$$K_{1s2s}=\frac{1}{4\pi\varepsilon_0}\frac{e^2}{2\pi^2}\int\frac{d\vec{k}}{k^2}g_1\left(\vec{k}\right)g_2\left(\vec{k}\right), \tag{9.100}$$

where

$$g_1\left(\vec{k}\right) = \int d\vec{r} e^{i\vec{k}\cdot\vec{r}} |\phi_{1s}(r)|^2$$
$$= \int_0^{+\infty} dr r^2 2\pi \int_0^{\pi} d\theta \sin\theta e^{ikr\cos\theta} |\phi_{1s}(r)|^2$$
$$= \frac{16}{\left(4 + a_0^2 k^2\right)^2} \tag{9.101}$$

and

$$g_2\left(\vec{k}\right) = \int d\vec{r} e^{-i\vec{k}\cdot\vec{r}} |\phi_{2s}(r)|^2$$
$$= \int_0^{+\infty} dr r^2 2\pi \int_0^{\pi} d\theta \sin\theta e^{ikr\cos\theta} |\phi_{2s}(r)|^2$$
$$= \frac{1 - 3a_{He}^2 k^2 + 2a_{He}^4 k^4}{\left(1 + 2a_{He}^4 k^4\right)^2}. \tag{9.102}$$

Therefore

$$K_{1s2s} = \frac{1}{4\pi\varepsilon_0} \frac{e^2}{2\pi^2} \int \frac{d\vec{k}}{k^2} \left(\frac{16}{\left(4 + a_0^2 k^2\right)^2}\right) \left(\frac{1 - 3a_{He}^2 k^2 + 2a_{He}^4 k^4}{\left(1 + 2a_{He}^4 k^4\right)^2}\right)$$
$$= \frac{17e^2}{162a_0} \pi\varepsilon_0. \tag{9.103}$$

The exchange energy is given by

$$J_{1s2s} = \frac{1}{4\pi\varepsilon_0} \frac{e^2}{2\pi^2} \int \frac{d\vec{k}}{k^2} \left(\int d\vec{r}_1 e^{i\vec{k}\cdot\vec{r}_1} \phi_{1s}(r_1) \phi_{2s}^{\star}(r_1)\right)$$
$$\times \left(\int d\vec{r}_2 e^{-i\vec{k}\cdot\vec{r}_2} \phi_{2s}(r_2) \phi_{1s}^{\star}(r_2)\right). \tag{9.104}$$

Using the intermediate result,

$$g_{12}\left(\vec{k}\right) = \int d\vec{r} e^{i\vec{k}\cdot\vec{r}} \phi_{1s}(r) \phi_{2s}^{\star}(r)$$
$$= \frac{256\sqrt{2} a_{He}^2 k^2}{\left(9 + 4a_{He}^2 k^2\right)^3}, \tag{9.105}$$

we finally get

$$J_{1s2s} = \frac{8e^2}{729 a_0 \pi\varepsilon_0}. \tag{9.106}$$

9.6 References

[1] M. Karplus and R. N. Porter, *Atoms and Molecules*, (The Benjamin Cummings Publishing Company, Menlo Park, CA, 1970).

[2] W. Heitler and F. London, Z. Phys. **44**, 455 (1927).

[3] D. Loss and D.P. DiVincenzo, "Quantum computation with quantum dots", Phys. Rev. A, **57**, 120 (1998).

[4] M.J. Calderón, B. Keiller and S. Das Sarma, "Exchange coupling in semiconductor nanostructures: validity and limitations of the Heitler-London approach", Phys. Rev. B, **74**, 045310 (2006).

[5] D. Melnikov, J-P Leburton, A. Taha and N. Sobh, "Coulomb localization and exchange modulation in two electron coupled quantum dots", Phys. Rev. B, **74**, 041309(R) (2006); D. V. Melnikov and J-P Leburton, "Dimensionality effects in the two-electron system in circular and elliptic quantum dots", Phys. Rev. B, **73**, 085320 (2006).

[6] F. Bloch, "Remarks on the electron theory of ferromagnetism and of electrical conductivity", Z. Physik, **57**, 545 (1929).

10

Spin Transport in Solids

One of the most important topics in spintronics is spin transport, i.e., how do spin polarized carriers travel through a piece of metal and semiconductor and what happens to their spin polarizations in the presence of various spin relaxing mechanisms such as D'yakonov-Perel', Elliott-Yafet, Bir-Aronov-Pikus and hyperfine interactions with nuclear spins. This topic has been dealt with extensively in the literature leading to a hierarchy of *spin transport models*. The two primary models are the so-called *drift-diffusion model* and the more sophisticated *semi-classical model*. In this chapter, we briefly discuss these two models. There are of course other models that have been used in the literature, but these two are the most popular and have emerged as the mainstays of spintronic device simulation.

10.1 The drift-diffusion model

A rigorous derivation of the drift-diffusion equations that describe spin transport in a semiconductor or metal can be found in ref. [1] which derives this equation starting from the density matrix and Wigner distribution function. That rigorous derivation is quite complex and the interested reader is referred to ref. [1] for the details. Here, we first present the final results from that work, and follow that with a much simpler (easier to understand) derivation.

Let n_Σ be the density of carriers with spin Σ in a two-dimensional electron gas (2-DEG). The equation that describes the spatio-temporal evolution of n_Σ in a driving electric field -E was shown in ref. [1] to be

$$\frac{\partial n_\Sigma}{\partial t} - \mathbf{D_s}\frac{\partial^2 n_\Sigma}{\partial x^2} - \mathbf{A}\frac{\partial n_\Sigma}{\partial x} + \mathbf{B}n_\Sigma = 0, \tag{10.1}$$

where

$$\mathbf{D_s} = \begin{pmatrix} D_s & 0 & 0 \\ 0 & D_s & 0 \\ 0 & 0 & D_s \end{pmatrix}, \tag{10.2}$$

$$\mathbf{A} = \begin{pmatrix} \mu_s E & 2\beta_{xz} D_s & 0 \\ -2\beta_{xz} D_s & \mu_s E & 0 \\ 0 & 0 & \mu_s E \end{pmatrix}, \tag{10.3}$$

$$\mathbf{B} = \begin{pmatrix} D_s\left(\beta_{xz}^2 + \beta_{yz}^2\right) & -\mu_s E\beta_{xz} & -\beta_{yx}\beta_{yz}D_s \\ \mu_s E\beta_{xz} & D_s\left(\beta_{xz}^2 + \beta_{yz}^2 + \beta_{yx}^2\right) & 0 \\ -\beta_{yx}\beta_{yz}D_s & 0 & D_s\beta_{yx}^2 \end{pmatrix}, \tag{10.4}$$

$$D_s = \frac{kT\tau_s}{m^*} \quad \mu_s = -\frac{e\tau_s}{m^*}. \tag{10.5}$$

The quantities β_{lk} are determined by spin-orbit interaction strengths, and τ_s is the spin relaxation time (determined by, among other things, the spin-orbit interaction strengths). Note that the spin diffusion coefficient D_s need not be equal to the charge diffusion coefficient D and the spin mobility μ_s need not be equal to the charge mobility μ since the spin relaxation time τ_s is generally not equal to the momentum relaxation time of charge carriers τ_m. Unfortunately, many published work in the spintronics literature has made the tacit assumption that $D = D_s$ and $\mu = \mu_s$. This is, of course, wrong and any conclusions based on this assumption are of dubious validity.

Ref. [1] has solved Equation (10.1) for the special case of spin injection in a two-dimensional electron gas residing in the x-y plane. Electrons are injected at $x = 0$. We reproduce those solutions here since we wish to highlight the difference between upstream transport (when carriers travel against the force exerted on them by the driving electric field) and downstream transport (carriers traveling along the force). The solutions of Equation (10.1) for this special case were found in ref. [1] to be:

$$\begin{aligned} n_{\Sigma_x}(x) &= n_{\Sigma_x}(0)exp\left[-\left(\frac{\mu E}{2D_s} + \sqrt{\left(\frac{\mu E}{2D_s}\right)^2 + B_{yz}^2}\right)x\right] cos\left(B_{xz}x\right), \\ n_{\Sigma_y}(x) &= n_{\Sigma_y}(0)exp\left[-\left(\frac{\mu E}{2D_s} + \sqrt{\left(\frac{\mu E}{2D_s}\right)^2 + B_{yz}^2}\right)x\right] sin\left(B_{xz}x\right), \\ n_{\Sigma_z}(x) &= n_{\Sigma_z}(0)exp\left[-\left(\frac{\mu E}{2D_s} + \left|\frac{\mu E}{2D_s}\right|\right)x\right]. \end{aligned} \tag{10.6}$$

The above equation shows that the ensemble averaged x-component of the spin oscillates with distance x with an exponentially decaying envelope $exp(-x/L_x)$. Similarly, the ensemble averaged y-component of the spin oscillates with distance x with an exponentially decaying envelope $exp(-x/L_y)$. The ensemble averaged z-component has no oscillatory component and decays exponentially with decay constant L_z. The decay constants L_x, L_y and L_z all depend on the driving electric field E. What is most interesting is that the decay constants depend on the *sign* of the electric field. If carriers are traveling "upstream" against the force exerted on them by the electric field, then the sign of E is positive and the decay lengths are much smaller than in the case when carriers are traveling "downstream" and the sign of E is negative. This means that spin polarization is randomized (or lost) much more rapidly in upstream transport compared to downstream transport. In fact,

the z-component of the spin does not decay at all (only in this special case) if transport is downstream (since $L_z \to \infty$), but decays rapidly if transport is upstream.

Equation (10.1), albeit rigorously derived, is not particularly user-friendly since it involves dyadics (9-component tensors that are 3×3 matrices). A simpler (and less rigorous) form of the one-dimensional drift diffusion equation that involves only scalars is [2]

$$\begin{aligned}\frac{\partial n_\uparrow}{\partial t} &= -\frac{n_\uparrow - n_\downarrow}{2\tau_{\uparrow\downarrow}} - \frac{\partial(n_\uparrow v_\uparrow)}{\partial x} + D_\uparrow \frac{\partial^2 n_\uparrow}{\partial x^2},\\ \frac{\partial n_\downarrow}{\partial t} &= -\frac{n_\downarrow - n_\uparrow}{2\tau_{\downarrow\uparrow}} - \frac{\partial(n_\downarrow v_\downarrow)}{\partial x} + D_\downarrow \frac{\partial^2 n_\downarrow}{\partial x^2},\end{aligned} \tag{10.7}$$

where the n-s are the concentrations, D-s are the diffusion coefficients and v-s are the drift velocities of up-spin and down-spin carriers. The quantity $1/\tau_{\uparrow\downarrow}$ is the rate at which up-spin electrons flip to down-spin, and $1/\tau_{\downarrow\uparrow}$ is the rate at which down-spin electrons flip to up-spin. In general, these two rates need not be equal.

The above equation pictures spin transport as transport of two coupled channels as shown in Figure 10.1. The up-spin and down-spin channels are coupled by spin flip processes described by the rates $1/\tau_{\uparrow\downarrow}$ and $1/\tau_{\downarrow\uparrow}$.

The disturbing feature of Equation (10.7) is that it deals with up-spin and down-spin electrons only. It really does not account correctly for superposition states which are superpositions of up-spin and down-spin. An electron can obviously exist in a superposition of up-spin and down-spin polarizations (e.g., if it is precessing about a magnetic field). Furthermore, the concept of spin flipping from up to down, or vice versa, cannot account for spin relaxation via mechanisms such as the D'yakonov-Perel' mode, since that mechanism is not a discrete spin flip event in time. It comes about because of asynchronous spin precession about the pseudo magnetic fields caused by spin orbit interactions. Such a process, which is continuous in time, cannot be described satisfactorily by a relaxation time involving discrete transitions between up-spin and down-spin states. The reader should bear in mind these issues when dealing with equations such as Equation (10.7).

Equation (10.7) is reminiscent of the standard drift-diffusion equations of charge transport [3] used to model bipolar transport (transport of both electrons and holes) in a semiconductor. The up-spin and down-spin electrons play the roles of electrons and holes in bipolar transport. The spin flip rates $1/\tau_{\uparrow\downarrow}$ and $1/\tau_{\downarrow\uparrow}$ act like electron-hole recombination rates. In the following subsection, we derive the steady-state version of Equation (10.7) by analogy with bipolar charge transport.

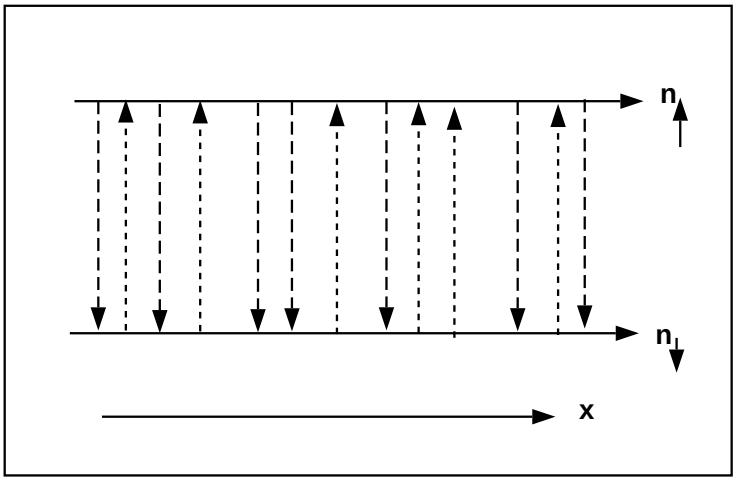


FIGURE 10.1
Pictorial depiction of the simplest spin drift-diffusion model. The up-spin and down-spin channels are coupled by spin flip processes. The long-dashed vertical arrows depict spin flip processes that convert an up-spin to down-spin and the short dashed vertical arrows depict the opposite process of flipping from down-spin to up-spin.

10.1.1 Derivation of the simplified steady state spin drift-diffusion equation

Analogous to the general current density expressions for electrons and holes, which include drift and diffusion components, the spin-polarized current densities of up-spin and down-spin populations are written as follows

$$J_{\uparrow} = n_{\uparrow} e\mu_{\uparrow} E + eD_{\uparrow} \frac{dn_{\uparrow}}{dx}, \tag{10.8}$$

and

$$J_{\downarrow} = n_{\downarrow} e\mu_{\downarrow} E + eD_{\downarrow} \frac{dn_{\downarrow}}{dx}, \tag{10.9}$$

where the first and second terms in each equation above are drift and diffusion components, respectively. The driving electric field is E, the electron charge is e and the mobilities of up-spin and down-spin populations are the $\mu_{\uparrow}$ and $\mu_{\downarrow}$, respectively.

The "continuity equations" for the up-spin and down-spin electron densities are given by

$$\frac{\partial n_{\uparrow}}{\partial t} = -\frac{n_{\uparrow}}{\tau_{\uparrow\downarrow}} + \frac{n_{\downarrow}}{\tau_{\downarrow\uparrow}} + \frac{1}{e}\frac{d}{dx} J_{\uparrow}, \tag{10.10}$$

$$\frac{\partial n_{\downarrow}}{\partial t} = -\frac{n_{\downarrow}}{\tau_{\downarrow\uparrow}} + \frac{n_{\uparrow}}{\tau_{\uparrow\downarrow}} + \frac{1}{e}\frac{d}{dx} J_{\downarrow}, \tag{10.11}$$

which includes in- and out-scattering rates for both populations, by analogy with electron-hole generation and recombination rates in the description of bipolar charge transport in semiconductors.

In steady state, the last two equations become

$$\frac{1}{e}\frac{dJ_{\uparrow}}{dx} - \frac{n_{\uparrow}}{\tau_{\uparrow\downarrow}} + \frac{n_{\downarrow}}{\tau_{\downarrow\uparrow}} = 0, \tag{10.12}$$

$$\frac{1}{e}\frac{dJ_{\downarrow}}{dx} - \frac{n_{\downarrow}}{\tau_{\downarrow\uparrow}} + \frac{n_{\uparrow}}{\tau_{\uparrow\downarrow}} = 0. \tag{10.13}$$

In the continuity equations, charge recombination processes are neglected since we consider unipolar spin transport.

Inserting the current density expressions (10.8) and (10.9) in these two last equations, we get

$$(\frac{d}{dx}\sigma_{\uparrow})E + \sigma_{\uparrow}(\frac{dE}{dx}) + eD_{\uparrow}\frac{d^2 n_{\uparrow}}{dx^2} = e(\frac{n_{\uparrow}}{\tau_{\uparrow\downarrow}} - \frac{n_{\downarrow}}{\tau_{\downarrow\uparrow}}), \tag{10.14}$$

and

$$(\frac{d}{dx}\sigma_{\downarrow})E + \sigma_{\downarrow}(\frac{dE}{dx}) + eD_{\downarrow}\frac{d^2 n_{\downarrow}}{dx^2} = e(\frac{n_{\downarrow}}{\tau_{\downarrow\uparrow}} - \frac{n_{\uparrow}}{\tau_{\uparrow\downarrow}}), \tag{10.15}$$

where $\sigma_{\uparrow(\downarrow)} = n_{\uparrow(\downarrow)} e\mu_{\uparrow(\downarrow)}$ are the up-spin(down-spin) conductivities.

Multiplying both sides of Equation (10.14) by $\sigma_\downarrow$ and both sides of (10.15) by $\sigma_\uparrow$ and subtracting, we get

$$X + Y = Z, \tag{10.16}$$

where

$$X = (\sigma_\downarrow \frac{d}{dx}\sigma_\uparrow - \sigma_\uparrow \frac{d}{dx}\sigma_\downarrow)E, \tag{10.17}$$

$$Y = eD_\uparrow\sigma_\downarrow \frac{d^2 n_\uparrow}{dx^2} - eD_\downarrow\sigma_\uparrow \frac{d^2 n_\downarrow}{dx^2}, \tag{10.18}$$

and

$$Z = en_\uparrow(\frac{\sigma_\downarrow}{\tau_{\uparrow\downarrow}} + \frac{\sigma_\uparrow}{\tau_{\uparrow\downarrow}}) - en_\downarrow(\frac{\sigma_\uparrow}{\tau_{\downarrow\uparrow}} + \frac{\sigma_\downarrow}{\tau_{\downarrow\uparrow}}). \tag{10.19}$$

With the use of the conductivity expressions above, the expression for X becomes

$$X = (e\sigma_\downarrow\mu_\uparrow \frac{dn_\uparrow}{dx} - e\sigma_\uparrow\mu_\downarrow \frac{dn_\downarrow}{dx})E. \tag{10.20}$$

For a homogeneous system, in which space-charge build up is neglected, any local electron charge due to $n_\uparrow(x)+n_\downarrow(x)$ must be balanced by a corresponding local hole charge. This does not preclude a net *spin polarization* of carriers, i.e., $n_\uparrow - n_\downarrow \neq 0$. In other words, spin polarization can be created without changing electron or hole densities. Charge conservation in unipolar materials dictates

$$\frac{d}{dt}(n_\uparrow + n_\downarrow) = 0, \tag{10.21}$$

but

$$\frac{d}{dt}(n_\uparrow - n_\downarrow) \neq 0. \tag{10.22}$$

If we rewrite the up-spin and down-spin electron densities as follows

$$n_\uparrow = (\frac{n_\uparrow + n_\downarrow}{2}) + (\frac{n_\uparrow - n_\downarrow}{2}), \tag{10.23}$$

and

$$n_\downarrow = (\frac{n_\uparrow + n_\downarrow}{2}) - (\frac{n_\uparrow - n_\downarrow}{2}), \tag{10.24}$$

then the expression for X becomes

$$X = \left[e\sigma_\downarrow\mu_\uparrow \frac{d}{dx}(\frac{n_\uparrow - n_\downarrow}{2}) + e\sigma_\uparrow\mu_\downarrow \frac{d}{dx}(\frac{n_\uparrow - n_\downarrow}{2})\right] E \tag{10.25}$$

or

$$X = \left[\frac{e}{2}(\sigma_\uparrow\mu_\uparrow + \sigma_\uparrow\mu_\downarrow)\frac{d}{dx}(n_\uparrow - n_\downarrow)\right] E. \tag{10.26}$$

The expression for Y is

$$Y = e(D_{\downarrow}\sigma_{\uparrow} + D_{\uparrow}\sigma_{\downarrow})\frac{d^2}{dx^2}(\frac{n_{\uparrow} - n_{\downarrow}}{2}), \tag{10.27}$$

while the expression for Z is

$$Z = e(\frac{n_{\uparrow} - n_{\downarrow}}{2})(\sigma_{\downarrow} + \sigma_{\uparrow})\frac{1}{\tau_{\uparrow\downarrow}} - e(\frac{n_{\downarrow} - n_{\uparrow}}{2})(\sigma_{\downarrow} + \sigma_{\uparrow})\frac{1}{\tau_{\downarrow\uparrow}} \tag{10.28}$$

or

$$Z = e(\frac{n_{\uparrow} - n_{\downarrow}}{2})(\sigma_{\uparrow} + \sigma_{\downarrow})(\frac{1}{\tau_{\uparrow\downarrow}} + \frac{1}{\tau_{\downarrow\uparrow}}). \tag{10.29}$$

Substituting the expressions for X, Y and Z in Equation (10.16), we get

$$\frac{eE}{2}(\sigma_{\uparrow}\mu_{\downarrow} + \sigma_{\downarrow}\mu_{\uparrow})\frac{d}{dx}(n_{\uparrow} - n_{\downarrow}) + \frac{e}{2}(D_{\uparrow}\sigma_{\downarrow} + D_{\downarrow}\sigma_{\uparrow})\frac{d^2}{dx^2}(n_{\uparrow} - n_{\downarrow}) = \frac{e}{2}(n_{\uparrow} - n_{\downarrow})(\sigma_{\uparrow} + \sigma_{\downarrow})\frac{1}{\tau_s}, \tag{10.30}$$

where we have introduced the average spin relaxation time

$$\frac{1}{\tau_s} = \frac{1}{\tau_{\uparrow\downarrow}} + \frac{1}{\tau_{\downarrow\uparrow}}. \tag{10.31}$$

By introducing the effective mobility

$$\mu_{eff} = \frac{\sigma_{\uparrow}\mu_{\downarrow} + \sigma_{\downarrow}\mu_{\uparrow}}{\sigma_{\uparrow} + \sigma_{\downarrow}} \tag{10.32}$$

and the effective diffusion constant

$$D_{eff} = \frac{\sigma_{\uparrow}D_{\downarrow} + \sigma_{\downarrow}D_{\uparrow}}{\sigma_{\uparrow} + \sigma_{\downarrow}} \tag{10.33}$$

and the intrinsic spin diffusion length

$$L_s = \sqrt{D_{eff}\tau_s}, \tag{10.34}$$

Equation (10.30) can be written in the compact form

$$\frac{d^2}{dx^2}(n_{\uparrow} - n_{\downarrow}) + \frac{eE\mu_{eff}}{eD_{eff}}\frac{d}{dx}(n_{\uparrow} - n_{\downarrow}) - \frac{(n_{\uparrow} - n_{\downarrow})}{L_s^2} = 0, \tag{10.35}$$

which is obviously the *steady state* version of Equation (10.7).

We search for general solutions to Equation (10.35) of the form

$$n_{\uparrow} - n_{\downarrow} = Ae^{-\lambda x} \tag{10.36}$$

corresponding to spin polarization decaying exponentially in space.

Substitution of this solution in Equation (10.35) yields

$$\lambda^2 - \lambda(\frac{\mu_{eff}E}{D_{eff}}) - \frac{1}{L_s^2} = 0 \tag{10.37}$$

or

$$\lambda = \frac{\mu_{eff}E}{2D_{eff}} \pm \sqrt{(\frac{\mu_{eff}e}{2D_{eff}})^2 + \frac{1}{L_s^2}}. \tag{10.38}$$

We want to avoid solutions which blow up at $\pm\infty$ since these are unstable and inadmissible. Suppose a continuous spin imbalance $(n_\uparrow - n_\downarrow)|_o$ is injected at x =0, and the electric field is along the -x direction. Since the spin polarization must gradually decay with distance from the point of injection, the distribution of the spin polarization is given by:

For $x > 0$

$$(n_\uparrow - n_\downarrow)(x) = (n_\uparrow - n_\downarrow)|_o e^{-\frac{x}{L_d}}, \tag{10.39}$$

where

$$L_d = \left[-\frac{|E|}{2}\frac{\mu_{eff}}{D_{eff}} + \sqrt{(\frac{|E|}{2}\frac{\mu_{eff}}{D_{eff}})^2 + \frac{1}{L_s^2}} \right]^{-1}. \tag{10.40}$$

For the upstream direction, i.e., in the direction of the electric field, we have for $x < 0$:

$$(n_\uparrow - n_\downarrow)(x) = (n_\uparrow - n_\downarrow)|_o e^{\frac{x}{L_u}}, \tag{10.41}$$

where

$$L_u = \left[\frac{|E|}{2}\frac{\mu_{eff}}{D_{eff}} + \sqrt{(\frac{|E|}{2}\frac{\mu_{eff}}{eD_{eff}})^2 + \frac{1}{L^2}} \right]^{-1}. \tag{10.42}$$

These expressions should be compared with the expressions in Equation (10.6).

Furthermore, we have

$$L_u L_d = L_s^2. \tag{10.43}$$

10.2 The semiclassical model

The drift-diffusion model just described has several intrinsic shortcomings. The first is that it is obviously incapable for describing any non-linear effects since the drift-diffusion equations are linear differential equations. Non-linearities in spin transport have been observed experimentally [4, 5] and

therefore are a reality. The second is that the drift-diffusion model cannot handle superposition of up-spin and down-spin states satisfactorily [6]. As a result, it cannot treat coherence effects arising from superposition of up-spin and down-spin states. Recently, these superpositions were treated in a Bloch equation approach [7] and a generalized drift-diffusion type approach derived from the Boltzmann transport equation [8]. These are beyond the normal drift-diffusion formalism and are not discussed in this book. The interested reader can peruse the pertinent references. Third, the drift-diffusion formalism cannot handle hot electron effects in spin transport. These are important in many devices, such as the Spin Valve Transistor discussed in Chapter 12. Finally, the normal drift-diffusion formalism is not capable of incorporating spin relaxation via the D'yakonov-Perel' mechanism (as stated before) since that mechanism does not involve discrete spin flip events in time and therefore cannot be captured by a simple rate such as $1/\tau_{\uparrow\downarrow}$ depicting flipping from up-spin state to down-spin state. The D'yakonov-Perel' (D-P) relaxation is a continuous process in time. The D-P relaxation also has complex connections with momentum relaxation since the pseudo-magnetic fields associated with spin-orbit interactions – which cause the D-P relaxation – are velocity dependent (recall Chapter 8). Therefore, such a process cannot be described by a constant relaxation rate. Any model capable of handling the D-P process must be able to treat the momentum dependence of the spin-orbit coupling (the primary cause of D-P relaxation) self-consistently.

In reality, the temporal evolution of spin and the temporal evolution of the momentum of an electron cannot be separated when there is transport. A moment's reflection will convince the reader that the D-P spin relaxation (or depolarization) rates must be functionals of the electron distribution function in momentum space which continuously evolves with time when an electric field is applied to drive transport and there is random carrier scattering. Thus, the D-P relaxation rate is not a constant but is a dynamic variable. To account for all this, one must self consistently solve the Boltzmann Transport Equation describing the spatio temporal evolution of the electron distribution function and the Liouville equation describing the spatio temporal evolution of the spin density matrix. The Liouville equation will contain a momentum dependent damping term; hence, the need to simultaneously solve the Boltzmann equation and the Liouville equation. Such situations are best treated by Monte Carlo simulation, which has been recently adopted by a number of groups to study spin transport in quasi-two-dimensional structures (see, for example, refs. [6, 9] among many).

10.2.1 Spin transport in a quantum wire: Monte Carlo simulation

In this subsection, we will provide the basics of Monte Carlo methods. Consider the quasi one-dimensional semiconductor structure of Fig. 8.4. An electric field $\mathcal{E}_{\S}$ is applied along the axis of the quantum wire to induce charge

flow. In addition, there is another transverse electric field E_y that causes Rashba spin-orbit interaction.

Following Saikin, et al. [6], we treat the spin using the standard spin density matrix:

$$\rho_\sigma(t) = \begin{bmatrix} \rho_{\uparrow\uparrow}(t) & \rho_{\uparrow\downarrow}(t) \\ \rho_{\downarrow\uparrow}(t) & \rho_{\downarrow\downarrow}(t) \end{bmatrix} \quad , \tag{10.44}$$

which is related to the spin polarization component as $S_n(t) = Tr(\sigma_n \rho_\sigma(t))$ $(n = x, y, z)$. Over a small time interval δt (much smaller than the mean time between collisions), we will assume that no scattering takes place and that the electron's momentum changes slowly enough (in other words $\mathcal{E}_\S \delta t$ is small enough) that transport can be described by a constant (time-independent) momentum or velocity within this time interval*. We take this velocity to be the average of the velocity at the beginning and end of the time interval:

$$v = (v_{initial} + v_{final})/2 = v_{initial} + (q\mathcal{E}_\S \delta t)/(2m^*). \tag{10.45}$$

During this interval, the spin density matrix undergoes a unitary evolution according to

$$\rho_\sigma(t + \delta t) = e^{-iH_{so}(k)\delta t/\hbar} \rho_\sigma(t) e^{iH_{so}(k)\delta t/\hbar} \quad , \tag{10.46}$$

where $H_{so}(k)$ is the wavevector dependent spin-orbit interaction Hamiltonian.

Equation (10.46) describes the coherent rotation or precession of the spin vector about an effective magnetic field (or pseudo magnetic field) $\vec{B}_{so}$ caused by spin-orbit interaction. This spin precession is described by the equation

$$\begin{aligned} \frac{d\vec{S}}{dt} &= \vec{\Omega} \times \vec{S}, \\ \vec{\Omega} &= \frac{g\mu_B \vec{B}_{so}}{\hbar}. \end{aligned} \tag{10.47}$$

The reader should be familiar with this notion because it was discussed in Chapter 8, and is the basis of D-P relaxation. The strength of the pseudo magnetic field is proportional to the electron's velocity (recall Equation (8.18)) which changes with time. Within a time interval δt, the strength of the pseudo-magnetic field is determined by the magnitude of the average velocity during that time interval. Note that during this time interval, the spin dynamics is coherent and there is no "relaxation" since the evolution is *unitary*. However, there are two agents that ultimately cause relaxation via the D-P mechanism. The first is the electric field $\mathcal{E}_\S$ that changes the "average velocity" from one time interval δt to the next, thereby changing the effective pseudo magnetic field that an electron sees. The second is stochastic

*Within such an interval, the electron's momentum changes only because of the electric field accelerating the electron.

scattering that changes the "average velocity" between two successive intervals (separated by a scattering event) *randomly*. These two causative agents produce a distribution of spin states that results in effective relaxation. The evolution of the spin polarization vector $\mathbf{S}$ ($= S_x\mathbf{u}_x + S_y\mathbf{u}_y + S_z\mathbf{u}_z$; where $\mathbf{u}_n$ is the unit vector along the n-direction) can be viewed as coherent motion (rotation) and relaxation/depolarization (reduction in magnitude). The coherent dynamics causes the spin to oscillate and the relaxation causes the amplitude of the oscillation to decay in time. This type of relaxation is the D'yakonov-Perel' relaxation.

In Monte Carlo models, one assumes that over the short time interval δt, the electron's velocity is time-invariant and given by the average velocity in Equation (10.45). Consequently, the spin precession frequencies Ω are constant and independent of time in the interval δt. Accordingly, the solution of Equation (10.47) for the spin components yields:

$$\begin{aligned}
S_x(t+\delta t) &= -\frac{\Omega_R}{\Omega_T}\left[S_y(t)sin(\Omega_T\delta t) + \left(\frac{\Omega_D}{\Omega_T}S_z(t) - \frac{\Omega_R}{\Omega_T}S_x(t)\right)cos(\Omega_T\delta t)\right] \\
&\quad + \left(\frac{\Omega_D}{\Omega_T}\right)^2 S_x(t) + \frac{\Omega_D\Omega_R}{\Omega_T^2}S_z(t), \\
S_y(t+\delta t) &= S_y(t)cos(\Omega_T\delta t) + \left(\frac{\Omega_R}{\Omega_T}S_x(t) - \frac{\Omega_D}{\Omega_T}S_z(t)\right)sin(\Omega_T\delta t), \\
S_z(t+\delta t) &= \frac{\Omega_D}{\Omega_T}\left[S_y(t)sin(\Omega_T\delta t) + \left(\frac{\Omega_D}{\Omega_T}S_z(t) - \frac{\Omega_R}{\Omega_T}S_x(t)\right)cos(\Omega_T\delta t)\right] \\
&\quad + \left(\frac{\Omega_R}{\Omega_T}\right)^2 S_z(t) + \frac{\Omega_D\Omega_R}{\Omega_T^2}S_x(t),
\end{aligned} \tag{10.48}$$

where Ω_R and Ω_D are given in Equation (8.44) for a quantum wire and $\Omega_T = \sqrt{\Omega_R^2 + \Omega_D^2}$. All Ω-s are calculated at the average value of the velocities in the interval δt given by Equation (10.45). It is easy to verify from the above equations that spin is conserved for every individual electron, i.e.,

$$S_x^2(t+\delta t) + S_y^2(t+\delta t) + S_z^2(t+\delta t) = S_x^2(t) + S_y^2(t) + S_z^2(t) = |\mathbf{S}| \tag{10.49}$$

10.2.2 Monte Carlo simulation

The average value of the velocity in an interval δt depends on the initial value of the wavevector $v_{initial}$ at the beginning of the interval. Intervals are chosen such that a scattering event can occur only at the beginning or end of an interval. The choice of scattering events and the velocity state after the event (i.e., the time evolution of the velocity) are found from a Monte Carlo solution of the Boltzmann Transport Equation in a quantum wire [10, 11, 12].

Equations (10.48) are solved directly in the Monte Carlo simulator. One typically uses a time step δt of few femtoseconds and a 1,000 - 10,000 electron ensemble to collect spin statistics. The Monte Carlo solution yields the spin

components $S_x(t_n)$, $S_y(t_n)$ and $S_z(t_n)$ of every one of the electrons at every sampled instant of time t_n. From this distribution, one calculates how the magnitude of the ensemble averaged spin vector $| < \mathbf{S} > |$ decays with time. The characteristic time constant of this decay is sometimes interpreted as the spin relaxation time associated with D'yakonov-Perel' relaxation.

10.2.3 Specific examples: temporal decay of spin polarization

In Fig. 10.2, we show the results of a specific Monte Carlo simulation. Here we study the time evolution of an ensemble of spins in a quantum wire of width 30 nm and thickness 4 nm (refer to Fig. 8.4). The electrons travel along the axis of the wire under the influence of a driving electric field, suffer random scattering events due to phonons, while all the time precessing about the pseudo magnetic fields due to spin-orbit interaction. The symmetry breaking electric field E_y inducing the Rashba effect was assumed to be 100 kV/cm and the material was GaAs. Spins were injected from the left contact ($x = 0$) at time $t = 0$, polarized in the x-direction (along the wire axis). The lattice temperature was 30 K. Electron scattering due to confined optical phonons and bulk acoustic phonons was considered.

Fig. 10.2 shows how the magnitude of the average (ensemble averaged over all electrons) spin vector $\mathbf{S}$ decays with time for four different values of the driving electric field $\mathcal{E}_§$ applied along the axis of the wire. This field drives transport. Only D-P relaxation is considered here since it is known to be dominant over Elliott-Yafet or any other mechanism in semiconductor structures at the temperatures considered.

As expected, the decay time decreases with increasing electric field. There are two contributing factors for this trend. First, the average velocity v given by Equation (10.45) increases with increasing electric field $\mathcal{E}_§$. Hence the spins precess more rapidly since Ω_T increases with increasing v. Now consider the random effect of scattering which results in different initial velocity $v_{initial}$ for different electrons in a given time interval. This results in a larger standard deviation in v at higher electric fields $\mathcal{E}_§$ (and hence different Ω-s, or different precession rates) for different electrons in the same time interval. Ensemble averaging over the electrons therefore causes the magnitude of the spin vector $\mathbf{S}$ to decay more rapidly, resulting in faster spin relaxation (or depolarization). The second factor that contributes to this trend is that the frequency of scattering itself increases with increasing electric field. Scattering randomizes the Ω-s since it randomizes v. This also results in faster relaxation at stronger electric fields.

At first glance, it may be troubling to assimilate the fact that the magnitude of the ensemble averaged $\mathbf{S}$ ($= | < \mathbf{S} > |$) can decay with time. One should remember that the magnitude of $\mathbf{S}$ ($= |\mathbf{S}|$) is conserved only for an *individual* electron as shown by Equation (10.49), but the magnitude is not invariant when we ensemble average over *many* electrons. It is this *ensemble averaging*

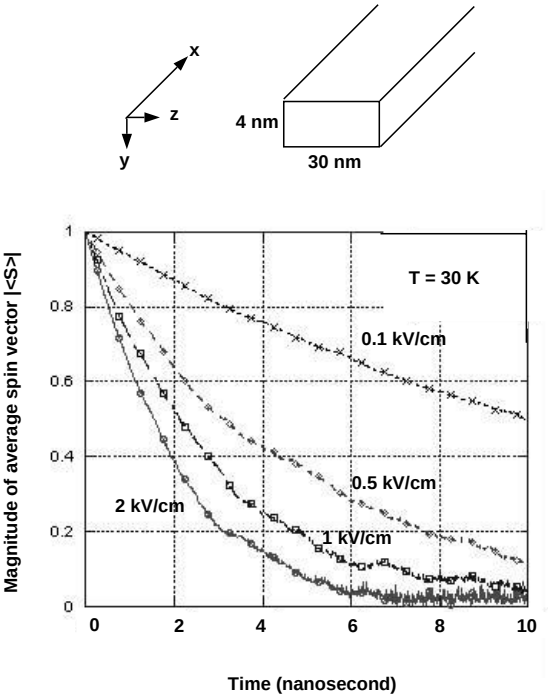


FIGURE 10.2
Temporal decay of the ensemble average spin vector with time in a GaAs quantum wire of dimension 4 nm × 30 nm at a lattice temperature of 30 K. The results are shown for various driving electric fields $\mathcal{E}_{\S}$. The spins are injected with their polarization initially aligned along the wire axis. At the top, we show the geometry of the quantum wire. Reprinted with permission from S. Pramanik, S. Bandyopadhyay and M. Cahay, Phys. Rev. B, **68**, 075313 (2003). Copyright (2003) American Physical Society. http://link.aps.org/abstract/PRB/v68/p075313.

that results in effective spin relaxation, as mentioned before.

10.2.3.1 Decay of spin components in time

In Fig. 10.3 we show how the average (ensemble averaged over all electrons) x-, y- and z-components of the spin decay with time. The driving electric field $\mathcal{E}_{\S} = 2$ kV/cm and the lattice temperature is 30 K. Since, initially the spin was polarized along the x-direction, the ensemble averaged y- and z-components remain near zero and the ensemble averaged x-component decays with time. The decay of the ensemble averaged x-component $< S_x >$ is indistinguishable from that of $| < \mathbf{S} > |$ in this case.

10.2.4 Specific examples: spatial decay of spin polarization

To study *spatial* decay of spin polarization (as opposed to *temporal* decay), Monte Carlo simulations are carried out where carriers are injected into a structure with a fixed spin polarization and then their spins are tracked in space as they transport under the influence of a driving electric field. We present here the example of a GaAs quantum wire of rectangular cross section 30 nm $\times$ 4 nm. We assume that the left contact is an ideal half metal magnetized along the x-direction that injects spins into the quantum wire with 100% efficiency. Therefore, electrons are injected into the quantum wire with injection velocities picked from a Fermi Dirac distribution, but all spins are aligned along the channel axis (x-axis) at $x = 0$ in order to simulate 100% efficient spin injection from the half metallic ferromagnetic contact.

In Fig. 10.4, we show the spatial decay of the ensemble averaged spin $| < S > |$ for two different temperatures of 30 K and 77 K. The spin decays faster at the higher temperature because of increased phonon scattering. If we (heuristically) define the spin relaxation length L_s as the distance it takes for $| < S > |$ to decay to $1/e$ times its initial value, then the relaxation length is quite large – greater than 10 μm at 77 K and 20 μm at 30 K.

10.2.5 Upstream transport

In Figs. 10.2, 10.3 and 10.4, we present results for *downstream* spin transport. Note that the decays of $| < S > |$ in both time and space are approximately exponential, so that the drift-diffusion model is at least qualitatively admissible, even if it is not quantitatively correct. We now present results for *upstream* transport and show that the drift-diffusion model completely breaks down in this case. For upstream motion, the decay of spin in space is not even monotonic, let alone being exponential. Such dynamics cannot be described even qualitatively by the drift-diffusion model.

In Fig. 10.5, we show the spatial evolution of the ensemble averaged spin $| < S > |$ of upstream electrons that were injected into a quantum wire of cross section 30 nm $\times$ 4 nm against the force exerted on them by a driving

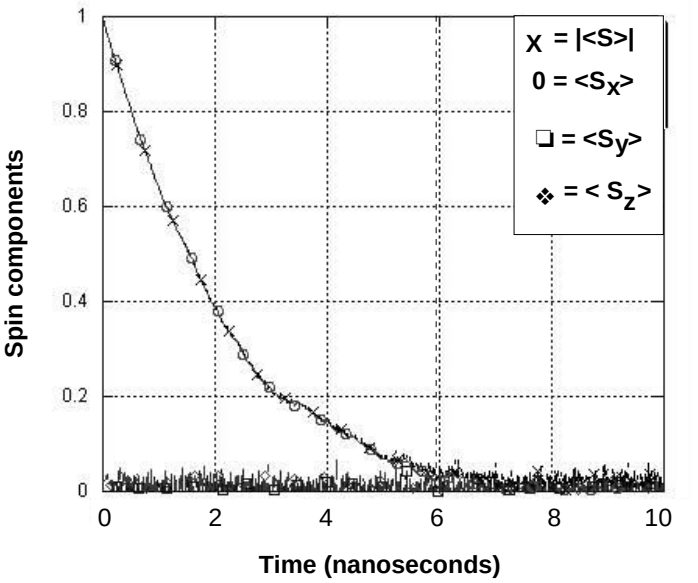


FIGURE 10.3
Temporal decay of the x-, y- and z-components of spin in the same GaAs wire at 30 K. The driving electric field is 2 kV/cm and the spins are injected with their polarization initially aligned along the wire axis (x-axis). The y- and z-components of the spin remain zero throughout. Reprinted with permission from S. Pramanik, S. Bandyopadhyay and M. Cahay, Phys. Rev. B, **68**, 075313 (2003). Copyright (2003) American Physical Society. http://link.aps.org/abstract/PRB/v68/p075313.

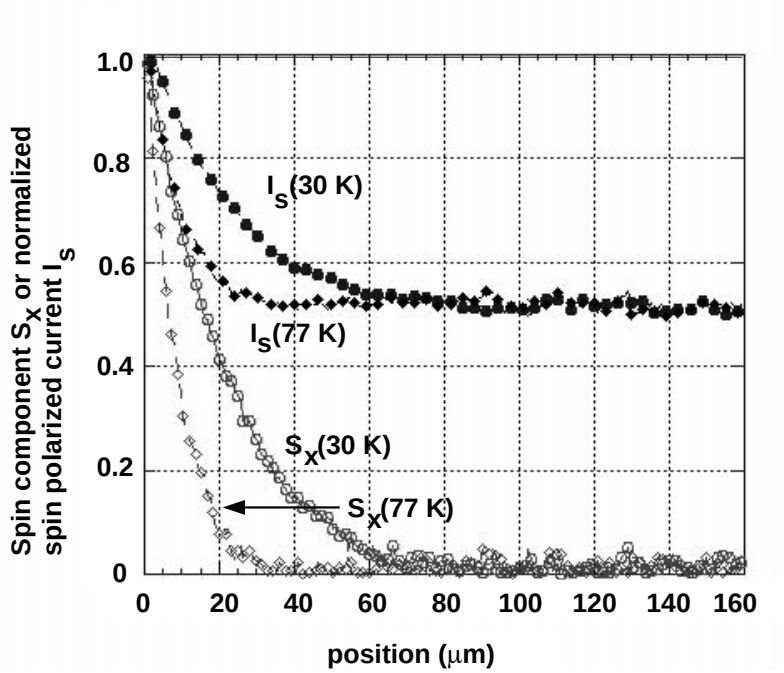


FIGURE 10.4
Spatial decay of the ensemble average spin vector in a GaAs quantum wire of cross-section 4 nm × 30 nm at lattice temperatures of 30 K and 77 K. The spins are injected into the quantum wire at $x = 0$ from a perfect spin injector with their polarization initially aligned along the wire axis. The driving electric field is 2 kV/cm. Also shown is the current I_s transmitted by a spin detector that is magnetized in the same direction (parallel) as the spin injector. When the spins are completely randomized, I_s will drop to 50% of its initial value. Reprinted with permission from S. Pramanik, S. Bandyopadhyay and M. Cahay, Applied Physics Letters, **84**, 266 (2004). Copyright (2004) American Institute of Physics.

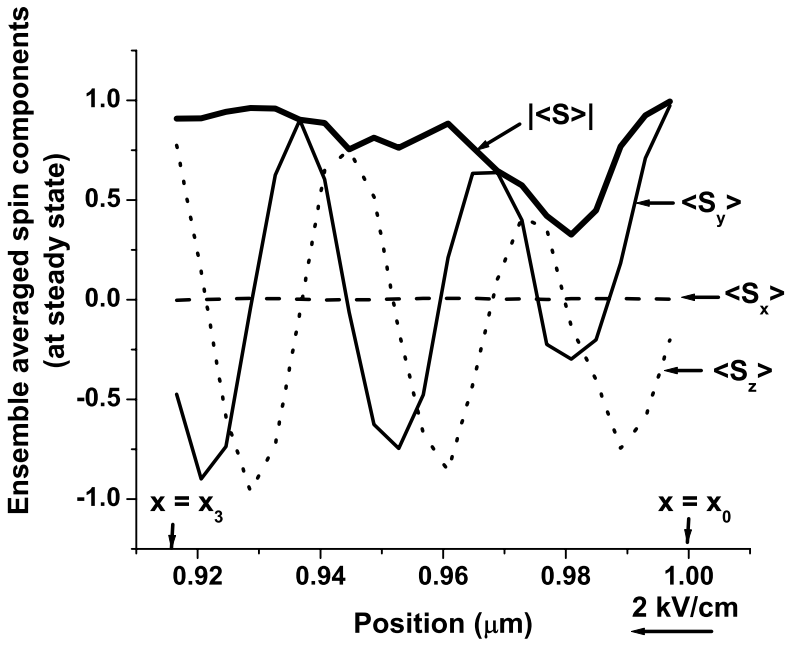


FIGURE 10.5

Spatial variation of ensemble averaged spin components for driving electric field $E_x = 2$kV/cm at steady state. Transport is upstream. Lattice temperature is 30 K and injection energy $E_0 = 426$ meV. Electrons are injected with velocities against the force exerted on them by the driving electric field. They are injected equally into the three lowest subbands, but ultimately end up in the highest subband (subband 3). This peculiar population inversion, which is very counter-intuitive and well beyond the drift-diffusion model, has been explained in Phys. Rev. B, **73**, 125309 (2006). The classical turning point of electrons traveling upstream in the highest subband (subband 3) is denoted by x_3 and x_{scat} indicates the point along the channel axis where subbands 1 and 2 get virtually depopulated. The classical turning point is the point where injected electrons have exhausted all the injection energy by traveling against the electric field, and therefore must turn around to travel in the direction of the force exerted by the electric field. A few electrons are found beyond the classical turning point since backscattering can push them beyond that point. This feature, too, is beyond the drift-diffusion picture. Injected electrons are $\hat{y}$ polarized and $x = x_0 = 1\ \mu m$ is the point of injection. Reprinted with permission from S. Pramanik, S. Bandyopadhyay and M. Cahay, Phys. Rev. B, **73**, 125309 (2006). Copyright (2006) American Physical Society. http://link.aps.org/abstract/PRB/v73/p125309.

electric field of 2 kV/cm. The injection energy is such that an injected electron can access the three lowest transverse subbands in the wire. Initially, each subband is populated equally, i.e., an injected electron has equal probability of going into any of the three subbands. A constant stream of injected electrons is maintained by continued injection to supplant those that exhaust their initial injection energies traveling upstream and finally turn around to become downstream travelers. The wire has Rashba and Dresselhaus interactions (exactly similar to the wire of Figs. 10.2, 10.3 or 10.4). We consider only D-P relaxation.

Note that at first, $| < S > |$ decays as electrons travel upstream. But then $| < S > |$ recovers, and finally climbs up to 1 and settles down there with no further change! Thus, we do not see a monotonic decay unlike in the case of downstream transport. What has happened is that two out of the three subbands have been depopulated of upstream electrons after they travel some distance. Surprisingly, the lowest two subbands depopulate and all the upstream electrons end up in the highest subband. This strange "population inversion" that happens in upstream transport has been explained in ref. [13]. When a single subband is occupied in a quantum wire (in this case the highest subband), there can be no further D-P relaxation as was shown in Problem 8.3. That is the reason why $| < S > |$ saturates at 1. Of course, if we included the Elliott-Yafet mechanism, then there would have been some additional spin relaxation and $| < S > |$ would be less than unity. Nonetheless, since the D-P relaxation is dominant, we would still most likely have seen the non-monotonic spin evolution in space. This behavior is completely beyond the drift-diffusion model.

10.3 Concluding remarks

In this chapter, we have introduced two models that are used to describe spin transport in solids. The drift-diffusion model is simple, user friendly, but often inadequate and inaccurate, particularly in upstream transport, where it breaks down completely. A more advanced model is the semiclassical formalism based on Monte Carlo simulation which captures many of the subtle features that are beyond the drift-diffusion model. However, it is computationally much more demanding. The Monte Carlo technique has currently (at the time of writing this book) become quite popular and is widely used for spintronic device simulation (see, for example, [6]). There are, of course, other models for simulating spin transport. This is a rapidly developing field and new models, or old models with new twists, appear in the literature frequently. The reader may scour the literature to find the other relevant models that have been proposed and developed to describe propagation of spins in solids.

They all have their strengths and weaknesses.

10.4 Problems

- **Problem 10.1**

  Consider a unipolar material and neglect space charge build up. Furthermore assume that $\tau_{\uparrow\downarrow} = \tau_{\downarrow\uparrow}$ and $D_\uparrow = D_\downarrow$. Show using the simple drift-diffusion equations that the total drift current $J_{drift} = n_\uparrow v_\uparrow + n_\downarrow v_\downarrow$ is spatially invariant, i.e., $\partial J_{drift}/\partial x = 0$.

- **Problem 10.2** Consider a spatially homogeneous system in which $\tau_{\uparrow\downarrow} = \tau_{\downarrow\uparrow}$. Using the drift-diffusion equation (10.7), show that the spin polarization density $n_\uparrow - n_\downarrow$ cannot change with time.

10.5 References

[1] S. Saikin, "A drift diffusion model for spin polarized transport in a two-dimensional non-degenerate electron gas controlled by spin-orbit interaction", J. Phys: Condens. Matt., **16**, 5-71 (2004).

[2] Y. Qi, Z. G. Yu and M. E. Flatte, "Spin Gunn Effect", Phys. Rev. Lett., **96**, 026602 (2006); Z. G. Yu and M. E. Flatté, "Electric field dependent spin diffusion and spin injection into semiconductors", Phys. Rev. B, **66**, 201202 (2002).

[3] R. F. Pierret, *Semiconductor Fundamentals*, Vol. 1, Purdue Modular Series on Solid State Devices, 2nd edition, (Addison-Wesley, Reading, MA, 1989).

[4] G. Schmidt, C. Gould, P. Grabs, A. M. Lunde, G. Richter, A. Slobodsky and L. W. Molenkamp, "Spin injection in the non-linear regime: band bending effects", www.arXiv.org/cond-mat/0206347.

[5] H. Sanada, Y. Arata, Y. Ohno, Z. Chen, K. Kayanuma, Y. Oka, F. Matsukura and H. Ohno, "Relaxation of photo-injected spins during drift transport in GaAs", Appl. Phys. Lett., **81**, 2788 (2002).

[6] S. Saikin, M. Shen, M. C. Cheng and V. Privman, "Semiclassical model for in-plane transport of spin polarized electrons in III-V heterostructures", www.arXiv.org/cond-mat/0212610; M. Chen, S. Saikin, M. C.

Cheng and V. Privman, "Monte Carlo modeling of spin polarized transport", www.arXiv.org/cond-mat/0302395; "Monte Carlo modeling of Spin FETs controlled by spin-orbit interaction", www.arXiv.org/cond-mat/0309118; M. Shen, S. Saikin and M. C. Cheng, "Monte Carlo modeling of spin injection through a Schottky barrier and spin transport in a semiconductor quantum well", www.arXiv.org/cond-mat/0405270.

[7] M. Q. Weng and M. W. Wu, "Kinetic theory of spin transport in n-type semiconductor quantum wells", J. Appl. Phys., **93**, 410 (2003).

[8] Y. Qi and S. Zhang, "Spin diffusion at finite electric and magnetic fields", Phys. Rev. B., **67**, 052407 (2003).

[9] A. Bournel, P. Dollfus, S. Galdin, F-X Musalem and P. Hesto, "Modelling of gate induced spin precession in a striped channel high electron mobility transistor", Solid State Commun., **104**, 85 (1997); A. Bournel, V. Delmouly, P. Dollfus, G. Tremblay and P. Hesto, "Theoretical and experimental considerations on the spin field effect transistor", Physica E, **10**, 86 (2001).

[10] D. Jovanovich and J-P Leburton, in *Monte Carlo Device Simulation: Full Band and Beyond*, Ed. K. Hess, (Kluwer Academic, Boston, 1991) pp. 191-218.

[11] N. Telang and S. Bandyopadhyay, "Effects of a magnetic field on hot electron transport in quantum wires", Appl. Phys. Lett., **66**, 1623 (1995).

[12] N. Telang and S. Bandyopadhyay, "Hot electron magnetotransport in quantum wires", Phys. Rev. B, **51**, 9728 (1995).

[13] S. Pramanik, S. Bandyopadhyay and M. Cahay, "Spin relaxation of upstream electrons in a quantum wire: Failure of the drift-diffusion models", Phys. Rev. B, **73**, 125309 (2006).

11

Passive Spintronic Devices and Related Concepts

In this chapter, we introduce the reader to the physics of important *passive* spin-based devices such as sensors and read heads for reading tiny magnetic fields associated with magnetically stored data. *Active* spin based devices (such as spin transistors) are discussed in the next chapter. The reader is introduced to the concepts of "spin valve", "giant magnetoresistance", and the notion of "spin injection efficiency" which affects the performance of almost every spintronic device extant. This chapter will also familiarize the reader with frequently found terminologies and show that several ideas that may appear unrelated at first are actually interlinked and could be understood in terms of a few simple principles.

11.1 Spin valve

The "spin valve" is a device that is widely used to study and measure properties associated with spin transport in paramagnetic metals or semiconductors. One important parameter that can be experimentally measured with a spin valve device is the *spin diffusion length* L_s that we discussed in Chapter 10. Knowledge of this parameter and the spin diffusion coefficient D_s yields the spin relaxation time τ_s based on the relation $L_s = \sqrt{D_s \tau_s}$. The spin valve is therefore an important measurement device. It is also at the heart of addressable data storage devices, such as magnetic random access memory (MRAM).

The basic spin valve device is simply a trilayered structure consisting of a paramagnetic layer sandwiched between two ferromagnetic layers. We describe several models to understand spin transport in this three-layer system.

First, a simple two resistor model, based on the two current model of charge transport in ferromagnets, is introduced for the ferromagnetic layers. This model was originally proposed by Mott [1]. Next, we discuss the Stoner-Wohlfarth model for the band structure of ferromagnets. This will allow us to determine the relative magnitudes of the two currents in the Mott model and is a simple extension of the Sommerfeld model of metals [2]. Finally, we extend the two current model to account for the presence of a paramagnetic

layer between the two ferromagnetic layers. Putting all these together, we can come up with a tractable model to describe the behavior of a "spin valve". A slightly more advanced treatment is then presented based on the transfer matrix formalism described in Chapter 15. The transfer matrix method allows us to include the effects of simple spin-independent and spin-dependent scattering mechanisms at the interfaces between the outer ferromagnetic layers (or contacts) and the inner paramagnetic layer, as well as scattering within the layers themselves. Such an analysis will highlight the highly sensitive nature of spin valve characteristics to the details of the contact's electron density of states (for a quick primer on the 'density of states', see Chapter 15), scattering at the interface between the contacts and the paramagnetic layer, and the potential barrier heights at the ferromagnet/paramagnet interfaces. Later, using the transfer matrix method, we will derive the *Jullière formula* used to relate the magnetoresistance of a spin valve to the spin polarizations of carriers at the Fermi energy in the contacts [3]. An extension of the Jullière formula can be used to relate the magnetoresistance (change of resistance in a magnetic field) of a spin valve structure to the spin diffusion length L_s in the paramagnetic layer, provided we assume that carriers enter the paramagnet via tunneling from the injecting ferromagnet and then drift/diffuse through the paramagnet with exponentially decaying spin polarization, before tunneling into the second (detecting) ferromagnetic layer.

The spin valve concept can be also applied to understand spin transport in multilayered structures composed of alternating layers of ferromagnetic and paramagnetic materials. This led to the discovery of the phenomenon of *giant magnetoresistance* (GMR) in the late 1980s [4, 5]. A simple resistor network theory of GMR is discussed in this chapter. The GMR effect is one of the early successes of spintronics and devices based on the GMR effect are used as "read head sensors" for reading data stored in magnetic storage media (computer hard disks, Apple iPods, etc.). The read heads are passive devices, but an active device, namely a "spin valve junction transistor", has also been proposed based on GMR in the base region. This is discussed in Chapter 12. The discovery of GMR has been recognized as an important advance in basic science by the 2007 Nobel Prize in Physics. Another application of spin valves is in magnetic random access memory (MRAM) which is a widely available commercial product based on spintronics. Suffice it to say then that the spin valve is an extremely important device, both from the perspective of fundamental spin physics and device applications.

11.2 Spin injection efficiency

Paramagnets do not intrinsically have any net spin polarization. In other words, the population of up-spin and down-spin electrons are equal under equilibrium. Only ferromagnets have a non-zero spin polarization of carriers under equilibrium conditions. Therefore, the only way that a net spin polarization Ξ given by

$$\Xi = \frac{n_\uparrow - n_\downarrow}{n_\uparrow + n_\downarrow} \tag{11.1}$$

(where $n_\uparrow$ is the concentration of up-spin electrons and $n_\downarrow$ is the concentration of down-spin electrons) can be generated in a paramagnet is either via the use of *electrical spin injection* of charge carriers from a ferromagnet or *optical spin injection*. The former (more relevant to electrical devices as opposed to optical and electro-optic devices) requires successfully injecting an imbalance of spin from a ferromagnet into the paramagnet in the form of a current.

This brings into focus the issue of *spin injection efficiency* at the ferromagnet/paramagnet interface. This is a measure of how efficiently a ferromagnet can inject spin into a paramagnet that it is in electrical contact with. A related quantity is the *spin detection efficiency*, which is a measure of how efficiently a ferromagnet can filter spin, i.e., how selectively it can transmit electrons of a particular spin orientation impinging on it from the paramagnet. If a ferromagnet injector injects only one kind of spin – either the majority or the minority spin exclusively– into a paramagnet, then it is an ideal spin injector and the spin injection efficiency is 100%. Similarly, if a ferromagnet detector transmits only one kind of spin, then the spin detection efficiency is 100%. Real ferromagnets will typically inject/transmit both majority and minority spins, albeit not equally. Therefore, spin injection and detection efficiencies tend to be less than 100%. Spin injectors are also sometimes referred to as "spin polarizers" in analogy with optics, and spin detectors are similarly referred to as "spin analyzers".

The spin-valve, GMR devices and the Spin Field Effect Transistors (discussed in the next chapter) all rely on high (ideally 100%) efficiency of electrical spin injection/detection across a ferromagnet/paramagnet interface for their operation. The spin injection efficiency is defined as

$$\zeta = \frac{J_\uparrow - J_\downarrow}{J_\uparrow + J_\downarrow}, \tag{11.2}$$

where $J_\uparrow$ is the current density due to up-spin electrons and $J_\downarrow$ is that due to down-spin electrons, at the interface. In the ferromagnet, $J_\uparrow \neq J_\downarrow$ since $n_\uparrow \neq n_\downarrow$. But can this inequality be sustained in the paramagnet as well? The current continuity equation, familiar to all students of device physics and engineering, mandates that in steady state, $\vec{\nabla} \cdot \vec{J} = 0$. Therefore,

$$J^\uparrow_{ferromagnet} + J^\downarrow_{ferromagnet} = J^\uparrow_{paramagnet} + J^\downarrow_{paramagnet}. \tag{11.3}$$

The above equation does not require that $J^{\uparrow}_{paramagnet} \neq J^{\downarrow}_{paramagnet}$. In other words, there is no guarantee that there will be a spin polarized current in the paramagnet. What would favor the inequality $J^{\uparrow}_{paramagnet} \neq J^{\downarrow}_{paramagnet}$ in the paramagnet (and therefore result in a spin polarized current) is "resistivity matching", i.e., the resistance of the ferromagnet and paramagnet should be about equal [6]. This would, of course, be impossible to attain when the ferromagnet happens to be a metal (e.g., cobalt or iron) and the paramagnet happens to be a semiconductor (e.g., GaAs or Si) or an insulator (e.g., Al_2O_3). Fortunately, there is a way out, which requires interposing a tunnel barrier between the ferromagnet and paramagnet [7]. This can circumvent the resistivity mismatch problem.

Much effort has been expended to improve spin injection efficiency at ferromagnet/paramagnet interfaces, culminating in the demonstration of $\zeta = 90\%$ at low temperatures [8] and 70% at room temperature [9]. Reference [8] actually used a semiconducting paramagnet that allowed good resistance matching and ref. [9] used a tunnel barrier with a metallic ferromagnet. These two strategies can result in relatively high spin injection efficiencies. However, for many applications, such as Spin Field Effect Transistors, even 90% is not nearly enough, and more than 99.999% is required for respectable performance as we will show in Chapter 12. This matter remains one of the major challenges in the entire field of spintronics.

11.2.1 Stoner-Wohlfarth model of a ferromagnet

It is imperative to have a basic understanding of the band structure of ferromagnets when attempting to understand such phenomena as spin injection and detection, as well as the Mott two-current model. The simplest band structure model of a ferromagnet is the so-called Stoner-Wohlfarth model. It was motivated by the two-current model of Mott [1] to describe spin transport in ferromagnets. Mott suggested that at temperatures well below the Curie temperature, most scattering mechanisms that an electron will encounter in a ferromagnet will preserve the direction of the electron spin. In other words, scattering will generally not cause a spin flip or spin relaxation. As a result, the majority and minority spin currents $J^{\uparrow}_{ferromagnet}$ and $J^{\downarrow}_{ferromagnet}$ can be thought as two independent components of the total current since no mixing occurs between them in the absence of spin flip scattering. The simplest way to implement this two-fluid or two-current model is to use an extension of the Sommerfeld model of metals [2] in which the energy-momentum relationship of both spin subbands is assumed to be parabolic and the mass of the carriers is the free electron mass*. The Stoner-Wohlfarth bandstructure (or dispersion relation) for a ferromagnet is depicted in Figure 11.1, in which the minimum of the conduction band for the minority spins is displaced upward

*It could be replaced by "effective mass", if needed.

(along the energy-axis) from the majority spin band by an amount equal to the exchange splitting energy Δ. The latter can be several electron-volts in many ferromagnets. The majority spins are those whose spin polarization is aligned parallel to the internal magnetic field in the ferromagnet[†], whereas the minority spins are the ones aligned opposite to this field. In this simple model, the $E - k$ relationship for majority spins is given by

$$E_{\uparrow} = \frac{\hbar^2}{2m_0}k^2, \tag{11.4}$$

whereas, for minority spins, it is equal to

$$E_{\downarrow} = \frac{\hbar^2}{2m_0}k^2 + \Delta. \tag{11.5}$$

At the Fermi energy, we define the Fermi wavevectors for the majority (up-spin) and minority (down-spin) electrons as

$$\hbar k_{\uparrow}^{F} = \sqrt{2m_0 E_F}, \tag{11.6}$$

and

$$\hbar k_{\downarrow}^{F} = \sqrt{2m_0(E_F - \Delta)}. \tag{11.7}$$

The band structure (or energy dispersion relations) in real ferromagnets is of course far more complicated than the one described by the Stoner-Wohlfarth model. In fact, for some of the most widely used ferromagnetic metals, the density-of-states at the Fermi level can be larger for the minority spin band than for the majority spin band. Two well known examples of this are cobalt and nickel. In a metal like cobalt, d-electrons are more numerous at the Fermi level, but there are also s-electrons. The d- and s-electrons tend to have anti-parallel spins. Using tunnel barriers of varying thickness, one can electrically inject either d- or s-electrons from cobalt into a paramagnet [10].

There is also a great deal of effort to create ferromagnets which are 100% spin polarized, i.e., for which the exchange splitting energy is larger than the Fermi energy measured from the bottom of the majority spin band. For such ferromagnets, the Fermi level will be placed below the bottom of the minority spin band, but above the bottom of the majority spin band. Assuming low enough temperatures such that the energy difference between the bottom of the minority spin band and the Fermi level is much larger than the thermal energy kT, we can assume that the minority band is completely unoccupied. In that case, we only have majority spins and the spin polarization Ξ given by Equation (11.1) is exactly 100%. Such materials are called ferromagnetic

[†]Within each 'domain' of the ferromagnet, there is an internal magnetic field which aligns spins either parallel or anti-parallel to it. In the unmagnetized state, the internal fields in different domains point in different directions. In the magnetized state, the field in every domain points in the direction of the magnetization.

half metals [11]. One well-known example of a half metal is lanthanum strontium manganate (LSMO) which has nearly 100% spin polarization at low temperatures. Some so-called Heusler alloys are also reportedly half metals.

Of course, perfectly ideal half metals may not actually exist. It has been argued that at any temperature above absolute zero, phonons and magnons will excite carriers to the minority spin band and destroy the 100% spin polarization [12]. Even at absolute zero, boundaries and inhomogeneities in a real sample will preclude 100% spin polarization. Nonetheless, the concept of half metals is important. In ideal half metals, $J^{\downarrow}_{ferromagnet} = 0$. Therefore, theoretically speaking, these materials should be capable of 100% efficient spin injection into a paramagnet. Insofar as 100% spin injection efficiency is an important target in spintronics, the study of half metals has received considerable attention from spintronics enthusiasts.

Exercises:

- In the Sommerfeld theory of a metal, the energy dependence of three-dimensional density of states is given by (see Chapter 15)

$$g_{3D}(E) = \frac{m_0}{\pi^2\hbar^3}\sqrt{2m_0E}. \tag{11.8}$$

Prove that, at zero temperature, $g_{3D}(E)$ can be rewritten as follows

$$g_{3D}(E) = \frac{3}{2}\frac{n}{E_F}\sqrt{\frac{E}{E_F}}, \tag{11.9}$$

where E_F and n are the Fermi energy and the electron density of the metal.

Solution: At zero temperature, the electron density of the metal is given by

$$n = \int_0^{E_F} dE g_{3D}(E). \tag{11.10}$$

Performing the integration using Equation (11.8),

$$n = \frac{\sqrt{2}m_0^{3/2}}{\pi^2\hbar^3}\frac{2}{3}E_F^{3/2}. \tag{11.11}$$

According to Equation (11.8),

$$g_{3D}(E) = \frac{\sqrt{2}m_0^{3/2}}{\pi^2\hbar^3}\frac{E_F^{3/2}}{E_F}\sqrt{\frac{E}{E_F}}, \tag{11.12}$$

which with the use of Equation (11.11) can be rewritten as

$$g_{3D}(E) = \frac{3}{2}\frac{n}{E_F}\sqrt{\frac{E}{E_F}}. \tag{11.13}$$

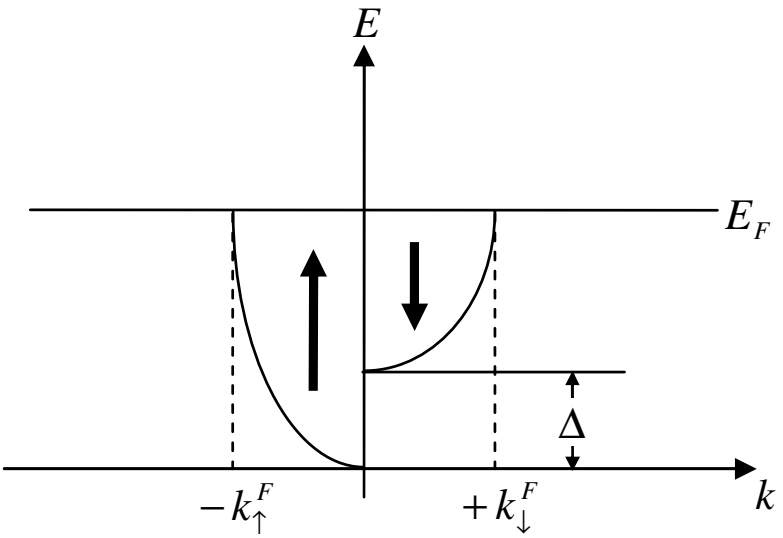


FIGURE 11.1
The Stoner-Wohlfarth model of a ferromagnet in which the energy dispersion relationships are assumed to be parabolic. The bottom of the minority spin subband is above that of the majority spin subband by an amount Δ (the exchange splitting energy). The effective mass of the carriers is assumed to be the free electron mass in both subbands. If the Fermi level is above the bottom of both spin subbands, then both subbands are occupied and carriers at the Fermi energy have a mixture of majority and minority spins. This is a normal ferromagnet. If the Fermi level is below the bottom of the minority spin subband but above that of the majority spin subband, then only the majority spin subband is occupied and the minority spin subband is unoccupied. Carriers with the Fermi energy have only majority spins and at absolute zero temperature, the spin polarization in the ferromagnet is ideally 100 % since all carriers have majority spin. This is an ideal half metallic ferromagnet.

- Using the results of the previous exercise, show that the density of majority spins $n_\uparrow$ is lower than the density of minority spins $n_\downarrow$. Calculate the ratio of $n_\uparrow/n_\downarrow$ in terms of E_F and Δ. Assume absolute zero temperature for simplicity.

- Show that if the Fermi level is well below the bottoms of both majority and minority spin subbands (the carrier populations are non-degenerate), so that Boltzmann statistics can replace Fermi Dirac statistics, then the spin polarization is given by

$$\Xi = \frac{n_\uparrow - n_\downarrow}{n_\uparrow + n_\downarrow} = \frac{e^{\Delta/kT} - 1}{e^{\Delta/kT} + 1} = tanh\left(\frac{\Delta}{2kT}\right). \tag{11.14}$$

11.2.2 A simple two resistor model to understand the spin valve

Consider two ferromagnetic contacts (F) in close proximity and separated by a narrow barrier of a paramagnetic material (N) as shown in Fig. 11.2. In the two fluid (or current) model of Mott, we can expect a difference between the resistance of the F/N/F junction when the magnetizations (indicated by arrows in the ferromagnetic layers) are either parallel (Fig.11.2(a)) or anti-parallel (Fig.11.2(b)). An electron whose spin is parallel to the magnetization of a ferromagnet will be a 'majority spin' in the ferromagnet and will travel more easily through that region. In other words, it encounters minimal resistance. On the other hand, an electron whose spin is anti-parallel to the magnetization will be a minority spin and will encounter greater resistance in traversing that region. All this can be understood from the Stoner-Wohlfarth band structure in Fig. 11.1. For a fixed energy, an electron with majority spin has higher velocity and therefore encounters less scattering from impurities and defects. Hence it encounters less resistance.

Based on this simple picture, we can estimate the resistance of the F/N/F junction using the resistor network model depicted in Fig. 11.2(c) and 11.2(d). For the case of parallel magnetization, the resistance for the up-spin (majority) band is given by $2R_\uparrow$ and the resistance of the down-spin (minority) band is given by $2R_\downarrow$ ($R_\downarrow >> R_\uparrow$). Now, in the Mott two-current model, the up-spin and down-spin channels are independent, so that the total resistance of the device is the parallel combination of the resistances $2R_\uparrow$ and $2R_\downarrow$. This is indicated in Fig. 11.2(c). As a result, the resistance R_P of the junction (here the subscript 'P' denotes the fact that the two ferromagnetic layers have parallel magnetizations) is given by

$$R_P = \frac{(2R_\uparrow)\,(2R_\downarrow)}{2R_\uparrow + 2R_\downarrow} = \frac{2R_\uparrow R_\downarrow}{R_\uparrow + R_\downarrow}. \tag{11.15}$$

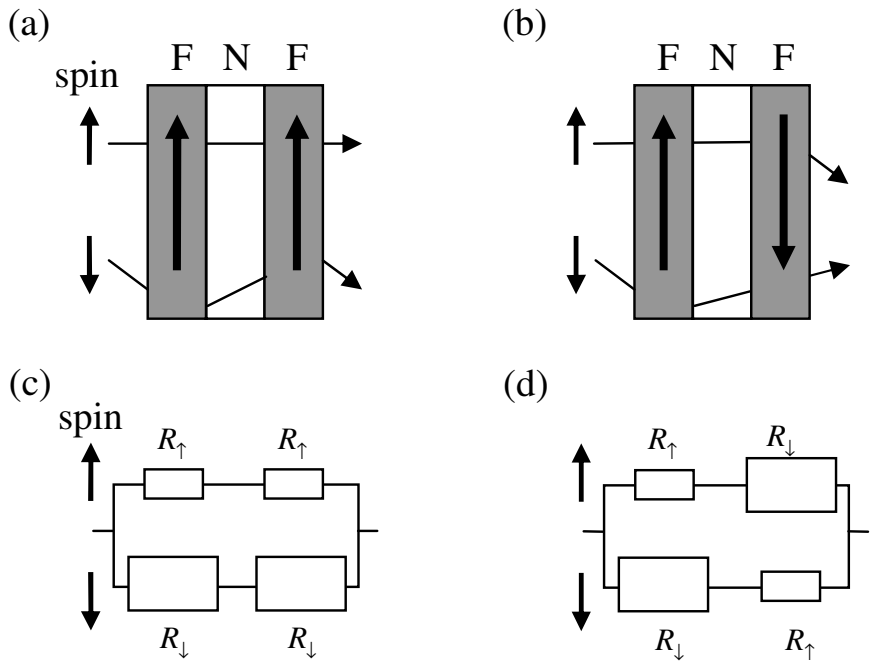


FIGURE 11.2
A F/N/F junction in which two ferromagnetic contacts (F) sandwich a paramagnetic (N) region a few nanometers thick. A spin valve effect comes from the difference of the overall resistance of the junction as the relative magnetization orientation of the two contacts is changed from parallel to anti-parallel. Figures (c) and (d) illustrate the simple two resistor network model of the junction and spin valve effect based on the two current model of the ferromagnetic contacts originally proposed by Mott [1]. Reproduced with permission from M. A. M. Gijs and G. E. W. Bauer, Adv. Phys., **46**, 285 (1997). Copyright (1997) Taylor & Francis.

When the two ferromagnetic layers have anti-parallel magnetizations, the two resistor model depicted in Fig. 11.2(d) leads to a resistance R_{AP} for the F/N/F junction given by

$$R_{AP} = \frac{R_\uparrow + R_\downarrow}{2}. \tag{11.16}$$

Therefore,

$$R_{AP} - R_P = \frac{(R_\downarrow - R_\uparrow)^2}{2\,(R_\downarrow + R_\uparrow)} > 0. \tag{11.17}$$

Consequently, the anti-parallel configuration has a higher resistance. Here we have assumed that the majority spins in one ferromagnet are also majority spins in the other. However, if that were not true, and the majority spin in one ferromagnetic layer is minority spin in the other (e.g., in the case of iron and cobalt as the two contacts), then the anti-parallel configuration may have the lower resistance.

The spin valve effect has been investigated for a wide variety of ferromagnetic contacts and intermediate paramagnetic layers. Three magnetoresistance ratios are commonly used to characterize the difference between the resistances R_P and R_{AP}:

(a) The tunneling magnetoresistance ratio

$$MR_1 = \frac{R_{AP} - R_P}{R_P} = \frac{G_P - G_{AP}}{G_{AP}}, \tag{11.18}$$

where $G_P = 1/R_P$ and $G_{AP} = 1/R_{AP}$ are the conductance of the junction in the parallel and anti-parallel configurations, respectively. This ratio is most often labeled using the acronym **TMR** in the literature.

(b) The junction magnetoresistance ratio

$$MR_2 = \frac{R_{AP} - R_P}{R_{AP}} = \frac{G_P - G_{AP}}{G_P}, \tag{11.19}$$

most often labeled with the acronym **JMR** in the literature, and finally,

(c) The spin conductance ratio

$$MR_3 = \frac{G_P - G_{AP}}{G_P + G_{AP}}. \tag{11.20}$$

These three ratios are related to each other.

Exercises:

- Establish the relationships between the different ratios MR_1, MR_2, MR_3 defined above, i.e., express MR_1 in terms of MR_2 or MR_3, and so on.

  Solution

  This is based on simple algebra and is left to the reader.

- Using the expressions for R_P and R_{AP} given above, express the ratios MR_1, MR_2, MR_3 in terms of the ratio $R_\uparrow/R_\downarrow$. Show that

$$MR_1 = \frac{\left(\frac{R_\uparrow}{R_\downarrow} - 1\right)^2}{4\frac{R_\uparrow}{R_\downarrow}},$$
$$MR_2 = \frac{\left(\frac{R_\uparrow}{R_\downarrow} - 1\right)^2}{\left(\frac{R_\uparrow}{R_\downarrow} + 1\right)^2},$$
$$MR_3 = \frac{\left(\frac{R_\uparrow}{R_\downarrow} - 1\right)^2}{\left(\frac{R_\uparrow}{R_\downarrow}\right)^2 + 6\frac{R_\uparrow}{R_\downarrow} + 1}. \tag{11.21}$$

11.2.3 More advanced treatment of the spin valve

The treatment in the previous section was based on the two-independent-current model of the ferromagnetic contacts due to Mott. No consideration was given to the paramagnetic layer in between. Obviously the paramagnetic layer must play a critical role since spin can relax in this layer causing mixing between the two channels (up-spin channel and down-spin channel), which will no longer remain independent. The quality of the interfaces between the paramagnet and the ferromagnets must also play a role in the magnetoresistance of the spin valve. Some (but not all) of these additional features are better described by the resistor network model shown in Fig. 11.3.

The resistance of a piece of metal with resistivity ρ, length l and cross-sectional area A is given by

$$R = \rho l/A. \tag{11.22}$$

In the Drude model for simple metals and semiconductors, ρ is given by

$$\rho = \frac{m^*}{ne^2\tau_m}, \tag{11.23}$$

where n is the electron density, m^* the electron (effective) mass, and τ_m the mean free time or average time between collisions for the carriers (actually, this is the 'momentum relaxation time' in the parlance of transport theory [13]).

In the two-current model of Mott, the spin-dependent resistance of a ferromagnet $R_{fm,\uparrow}$ and $R_{fm,\downarrow}$ can be assumed to be inversely proportional to the conductivity for the up-spin ($\sigma_\uparrow$) and down-spin channel ($\sigma_\downarrow$). We therefore have

$$R_{fm,\uparrow}/R_{fm,\downarrow} = \sigma_\downarrow/\sigma_\uparrow. \tag{11.24}$$

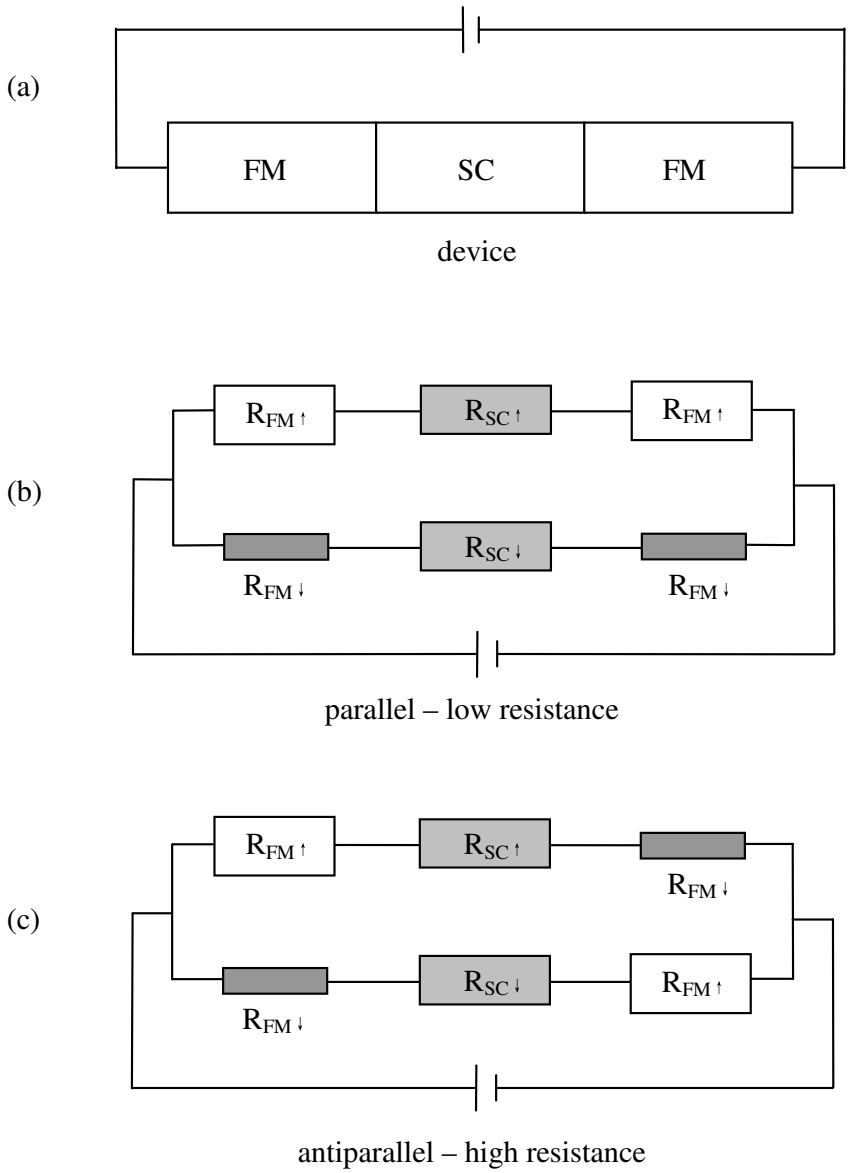


FIGURE 11.3
Advanced model of the spin valve effect taking into account the non-magnetic semiconductor sandwiched between two ferromagnetic contacts. Figures (b) and (c) show the resistor network model used to calculate the resistance of the spin valve for the case of parallel and anti-parallel magnetizations of the contacts, respectively. Reproduced with permission from G. Schmidt and L. W. Molenkamp, Semicond. Sci. Technol., **17**, 310 (2002).Copyright(2002) Institute of Physics.

Introducing a parameter β_p, which is the "conductivity spin polarization in the ferromagnet"[‡], given by

$$\beta_p = \frac{\sigma_\uparrow - \sigma_\downarrow}{\sigma_\uparrow + \sigma_\downarrow}, \tag{11.25}$$

we can then rewrite

$$R_{fm,\uparrow} = \frac{2R_{fm}}{1+\beta_p}, \tag{11.26}$$

$$R_{fm,\downarrow} = \frac{2R_{fm}}{1-\beta_p}, \tag{11.27}$$

where $R_{fm} = \frac{l}{(\sigma_\uparrow + \sigma_\downarrow)A}$.

If $\beta_p = 100\%$, i.e., only majority spins (up-spins) are present in the ferromagnet (half metallic ferromagnets), then $R_{fm,\downarrow} \to \infty$, indicating that the resistance encountered by (the absent) minority spins (down-spins) is infinity. This is consistent with the Mott picture.

A more realistic treatment of the spin valve effect must take into account the presence of the non-magnetic layer between the two contacts. The latter is assumed to be a paramagnetic semiconducting layer (SC) in Figure 11.3 and the same two current fluid model of Mott is used for the spin channels in the paramagnet as well. In the paramagnetic semiconductor, there are no "majority" or "minority" spins (both spins are equal in population). Therefore,

$$R_{sc,\uparrow} = R_{sc,\downarrow} = 2R_{sc}, \tag{11.28}$$

where R_{sc} is the resistance of the semiconducting channel which can be simply approximated by $R_{sc} = \rho_{sc} d/A$, ρ the resistivity of the semiconductor, d the separation between the contacts and A the cross-sectional area of the junction.

The resistance of the spin valve in the parallel and anti-parallel configurations can then be calculated using the resistor network model depicted in Figure 11.3(b) and 11.3(c).

Referring to Fig. 11.3(c), we see that, in the anti-parallel configuration, the resistances of both spin channels are equal and given by

$$R_\uparrow = R_\downarrow = R_{fm,\downarrow} + 2R_{sc} + R_{fm,\uparrow}. \tag{11.29}$$

As a result, the polarization of the spin current defined by the quantity

$$\alpha_p = \frac{R_\downarrow - R_\uparrow}{R_\downarrow + R_\uparrow} \tag{11.30}$$

[‡]Note that $\beta_p = \Xi$ if the mobilities of up-spin and down-spin electrons are equal, but not otherwise.

is equal to zero.

Referring to Fig. 11.3(b), the situation is different for the parallel case. Here, the polarization of the spin current is given by

$$\alpha_p = \frac{R_\downarrow - R_\uparrow}{R_\uparrow + R_\downarrow} = \frac{2(R_{fm,\uparrow} - R_{fm,\downarrow})}{2(R_{fm,\uparrow} + R_{fm,\downarrow}) + 4R_{sc}}. \tag{11.31}$$

This quantity is also sometimes referred to the *spin (current) injection efficiency*.

Using the expressions for $R_{fm,\uparrow}$ and $R_{fm,\downarrow}$ given above, this leads to

$$\alpha_p = \beta_p \left(\frac{R_{fm}}{R_{sc}}\right) \frac{2}{\left[\frac{2R_{fm}}{R_{sc}} + \left(1 - \beta_p^2\right)\right]}. \tag{11.32}$$

This last equation shows that the spin polarization of the current is proportional to the spin polarization β_p of the ferromagnetic contacts (for small β_p).

A plot of α_p versus β_p for different values of the ratio R_{sc}/R_{fm} is shown in Fig. 11.4. Obviously, the highest α_p is obtained when the spin polarization in the ferromagnets, β_p, is 100%. This shows that *half metals* are the best spin injectors and detectors, as we had surmised earlier.

In actual spin valves, the ratio R_{sc}/R_{fm} can be of the order of 10^4 leading to a very low spin polarization of the current in the semiconductor. It is therefore important that this ratio be close to unity if significant spin polarization in the semiconductor is desired. This is the basis of the infamous *resistance mismatch* problem pointed out in ref. [6] which insisted that in order to observe significant spin polarization in the semiconductor (i.e., a large value of α_p), the resistance of the semiconductor and ferromagnet should be "matched", i.e., $R_{sc} \approx R_{fm}$. Actually from Equation (11.32) it appears that if $R_{fm} >> R_{sc}$, then α_p approaches β_p. However, the ferromagnet is typically a metal and therefore unlikely to be more resistive than the semiconductor. Hence, this condition is impossible to meet with a metallic ferromagnet. This stimulated much research in semiconducting ferromagnets (or dilute magnetic semiconductor materials) until Rashba pointed out that the interposition of a tunnel barrier between the ferromagnet and the semiconductor can improve spin injection into the semiconductor [7] in spite of a resistance mismatch. However, tunnel barriers are inconvenient and reduce the overall conductance of the spin valve, diminishing its current-carrying capability. Consequently, research in dilute magnetic semiconductors has remained a big enterprise.

For a spin valve, the overall resistance in the parallel configuration is

$$R_P = \frac{(2R_{fm,\uparrow} + 2R_{sc})(2R_{fm,\downarrow} + 2R_{sc})}{2R_{fm,\uparrow} + 2R_{fm,\downarrow} + 4R_{sc}}, \tag{11.33}$$

while for the anti-parallel configuration, we get

$$R_{AP} = \frac{R_{fm,\uparrow} + R_{fm,\downarrow} + 2R_{sc}}{2}. \tag{11.34}$$

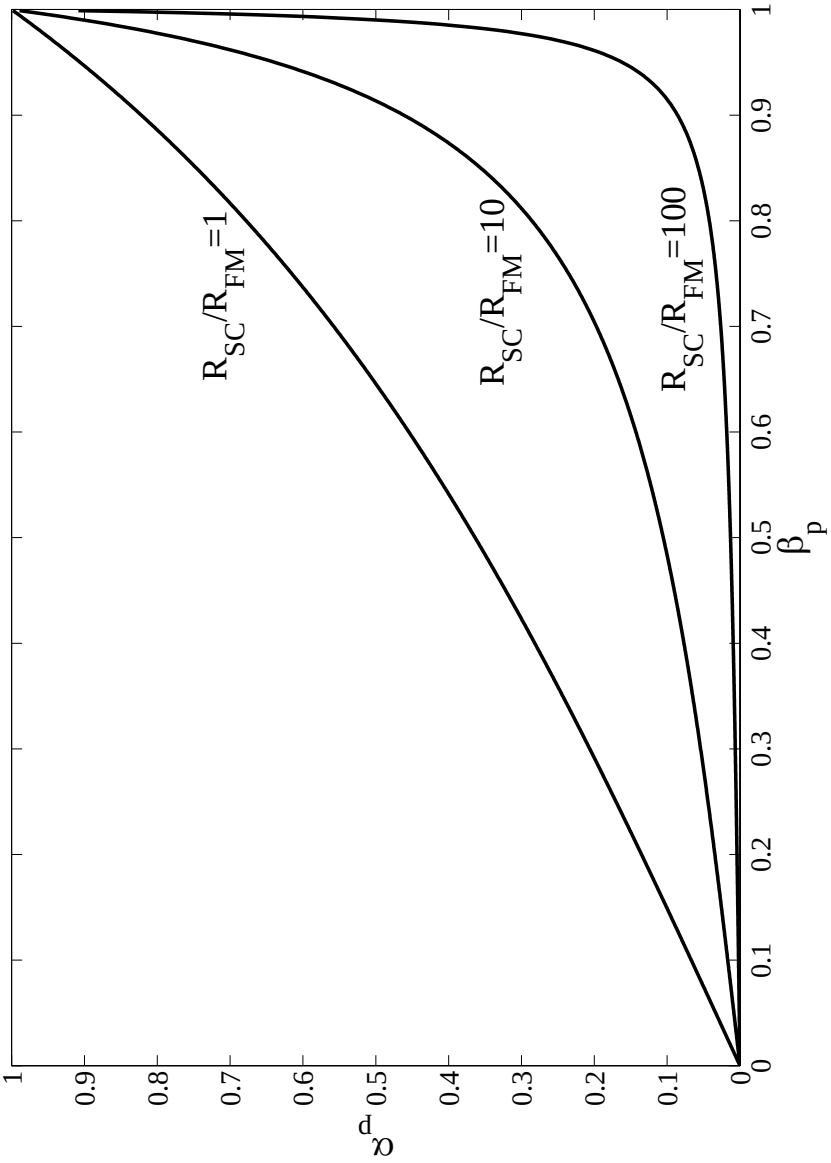


FIGURE 11.4
Plot of the spin current injection efficiency α_P versus the spin polarization β_P in the ferromagnetic contacts for different ratios of the resistances of the ferromagnetic contacts and semiconductor layer.

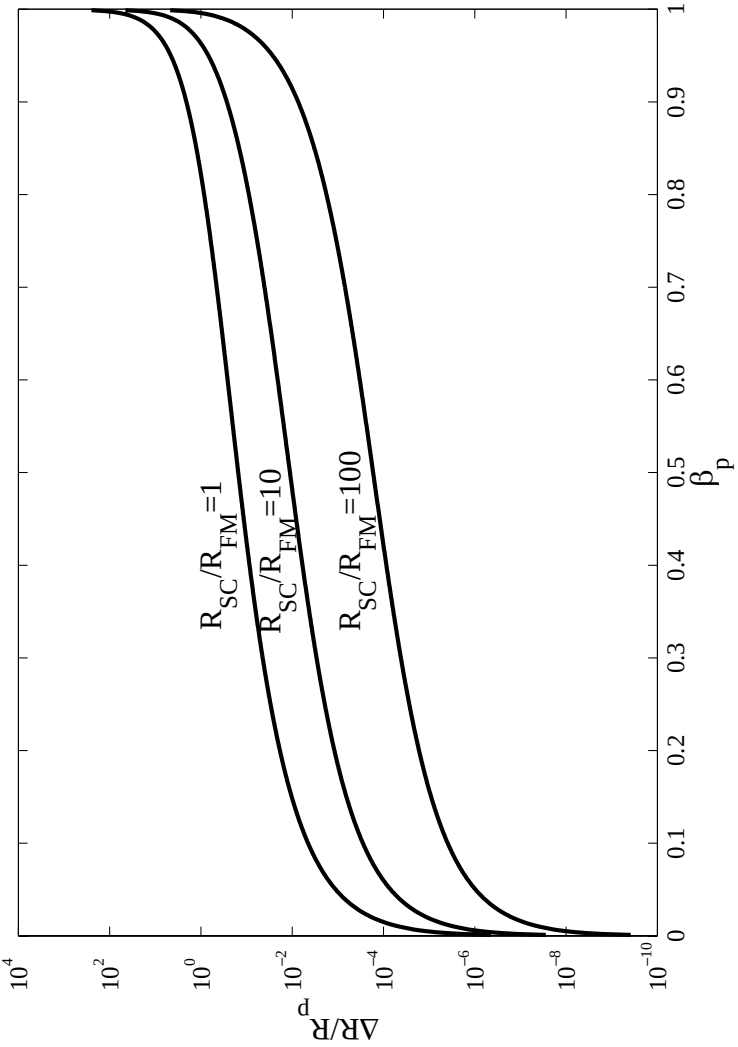


FIGURE 11.5
Plot of the tunnel magnetoresistance $\Delta R/R_p$ versus the spin polarization β_P of the ferromagnetic contacts for different ratios of the resistances of the ferromagnetic contacts and semiconductor channel.

Hence, the tunneling magnetoresistance ratio MR_1 defined earlier is given by

$$TMR = MR_1 = \frac{\Delta R}{R_P} = \frac{R_{AP} - R_P}{R_P} = \frac{\left(R_{fm,\uparrow} - R_{fm,\downarrow}\right)^2}{4\left(R_{fm,\uparrow} + R_{sc}\right)\left(R_{fm,\downarrow} + R_{sc}\right)}. \tag{11.35}$$

Using the values of $R_{fm,\uparrow}$ and $R_{fm,\downarrow}$ defined above, we get

$$MR_1 = \frac{\Delta R}{R_P} = \frac{\beta_p^2}{1-\beta_p^2}\left(\frac{R_{fm}}{R_{sc}}\right)^2 \left[\frac{4}{4\left(\frac{R_{fm}}{R_{sc}}\right)^2 + 4\left(\frac{R_{fm}}{R_{sc}}\right) + (1-\beta_p^2)}\right]. \tag{11.36}$$

Since MR_1 is proportional to R_{fm}^2/R_{sc}^2, it should be almost impossible to detect in an experiment if R_{fm}/R_{sc} is of the order of 10^{-4}. A plot of $MR_1 = \Delta R/R_P$ versus β_p for different values of the ratio R_{sc}/R_{fm} is shown in Fig. 11.5. Since β_p is of the order of 0.3-0.5 for normal metallic ferromagnets (*not* half metals) and R_{sc}/R_{fm} is typically of the order of 10^4, because of the large difference in the conductivities between non-magnetic semiconductor layers and metallic ferromagnetic contacts, $\Delta R/R_P$ is expected to be very small and therefore very difficult to detect in an experiment. Surely this could have had a chilling effect on research in spin valves. Fortunately, this did not happen.

The simple resistor network model seems to tell us that spin valves are doomed from the very start by the resistance mismatch problem. Fortunately, this is not the case and spin valves are now widely used as basic components of hard-drive read heads and magnetic random access memories. Therefore, clearly, the simple resistor network model does not always represent reality. There is a need for more advanced theoretical treatments of the spin valve effect, beyond the resistor network model, which we discuss in the next subsections.

Exercise:

- Using the resistor network model shown in Fig. 11.3, derive an expression of the other magnetoresistance ratios MR_2 and MR_3 in terms of β_p and the ratios R_{fm}/R_{sc}.
- Derive analytical expressions for α_p and MR_1 if the two ferromagnetic contacts are characterized by two different spin polarization factors β_{p1} and β_{p2}.
- If we define the parameter $\gamma = R_{fm}/R_{sc}$, show that

$$\alpha_p = \frac{2\beta_p\gamma}{2\gamma + 1 - \beta_p^2} \tag{11.37}$$

and

$$MR_1 = \frac{\Delta R}{R_P} = \frac{\beta_p^2}{1-\beta_p^2}\gamma^2 \left[\frac{4}{4\gamma^2 + 4\gamma + (1-\beta_p^2)}\right]. \tag{11.38}$$

Compute $d\alpha_p/d\gamma$ and $d\left(\Delta R/R_P\right)/d\gamma$ for $\beta_p = 0.5$ and plot those quantities as a function of γ.

11.2.4 A transfer matrix model

So far, the spin valve effect was modeled using a resistor network model, which did not include the effects of the interface between the contacts and the non-magnetic layer. In this section, we describe a simple one-dimensional model of the spin valve effect, in which the intermediate non-magnetic semiconductor layer is assumed to be so thin that it can be approximated as a delta scatterer, as shown in Figure 11.6. The scatterer has a spin-independent part of strength Γ and spin-dependent part of strength Γ'. Furthermore, we assume that the scattering potential has a spin-independent and spin-dependent part. The magnetoresistance of the spin valve is calculated using the transfer matrix formalism described in Chapter 15 based on a purely one-dimensional picture of spin transport. We assume that there is only one occupied subband in the one-dimensional (quantum wire) semiconductor layer. Furthermore, we assume absolute zero temperature and calculate the conductance of the majority and minority spin subbands using the Landauer conductance formula [14].

Case 1: Both Γ and Γ' are set equal to zero

In the parallel configuration, since there is no spin scattering at the interface, the conductance for both the majority and minority spins is given by

$$G_\uparrow = G_\downarrow = G_P, \tag{11.39}$$

and

$$G_P = 2e^2/h. \tag{11.40}$$

In the anti-parallel configuration, the majority spins incident from the left contact see a step of size Δ, the exchange splitting energy, as they transmit across the interface. For the minority spin band, incident from the left, this appears as a step down of size Δ, as shown in the middle panel of Fig. 11.6. The transmission coefficients for the majority and minority spin subbands are

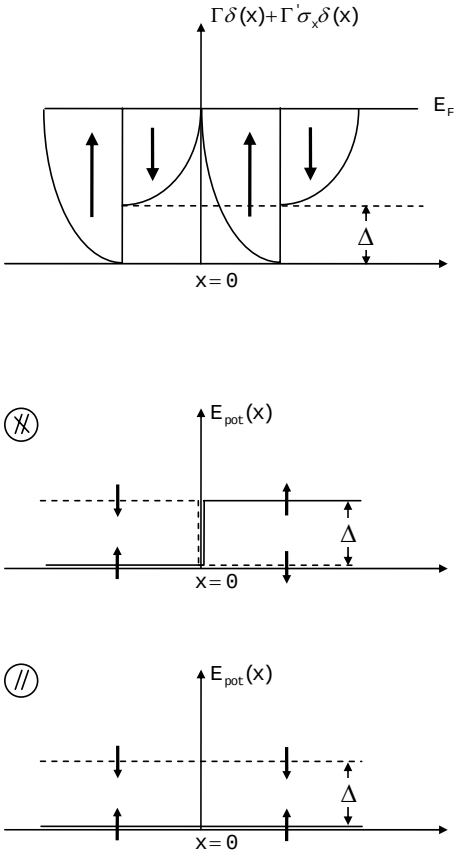


FIGURE 11.6
(Top panel) Simple one-dimensional model of a barrier between two identical ferromagnetic contacts. The scattering potential contains a spin-independent part of strength Γ and a spin-dependent part of strength Γ'. (Middle panel) In the absence of any scattering potential at the interface, when the magnetizations of the contacts are anti-parallel, the majority spins impinging from the left see a step barrier of height Δ (full line). The dashed line is the step down barrier for the minority spins incident from the left contact. (Bottom panel) When the magnetizations of both contacts are parallel, there is no potential step in the path of either spin flowing from one contact to the other.

identical and each is equal to

$$T_\uparrow = T_\downarrow = \frac{4k_\uparrow^F k_\downarrow^F}{\left[k_\uparrow^F + k_\downarrow^F\right]^2}. \tag{11.41}$$

As a result, the conductance of the spin valve in the anti-parallel configuration is given by

$$G_{AP} = \frac{8k_\uparrow^F k_\downarrow^F}{\left[k_\uparrow^F + k_\downarrow^F\right]^2}\frac{e^2}{h}. \tag{11.42}$$

A plot of the three different magnetoresistance ratios MR_1, MR_2 and MR_3 is shown in Figure 11.7 as a function of the ratio $\lambda = k_\uparrow^F/k_\downarrow^F$. This simple model reveals that the values of MR_1, MR_2, MR_3 can be quite different. Much more sophisticated treatments than the one described here also reveal this difference. Therefore, one must be very careful when comparing experimental data with theoretical models since all three definitions (MR_1, MR_2, MR_3) proliferate the literature.

Case 2: Γ and $\Gamma' \neq 0$

First, we calculate the conductance for majority spin incident from the left contact for which the incoming and outgoing components are depicted in Figure 11.8.

To calculate the reflection $(R_\uparrow, R_\downarrow)$ and transmission $(T_\uparrow, T_\downarrow)$ amplitudes, we must solve the Pauli Equation subject to the following boundary equations for the spinor at the interface

$$\begin{aligned} \psi(0_-) &= \psi(0_+) \\ \frac{d\psi}{dx}(0_+) &= \frac{d\psi}{dx}(0_-) + \frac{2m_0\Gamma}{\hbar^2}\left[1 + \left(\frac{\Gamma'}{\Gamma}\right)\sigma_x\right]\psi(0). \end{aligned} \tag{11.43}$$

Exercise: Starting with the Pauli Equation and the scattering potential shown in Fig. 11.6, prove that the boundary condition (11.43) must be satisfied by integrating the Pauli Equation across the interface.

Upon substituting the expressions for the spinors shown in Fig. 11.8, we get a set of four equations for the four unknowns $(R_\uparrow, R_\downarrow, T_\uparrow, T_\downarrow)$. This system of equations can be solved easily leading to

$$\begin{aligned} R_\uparrow &= \frac{-i\hbar^2 k_\uparrow^F m_0\Gamma + m_0^2\left(\Gamma^2 - \Gamma'^2\right)}{A}, \\ R_\downarrow &= \frac{-i\hbar^2 k_\uparrow^F m_0\Gamma'}{A}, \end{aligned}$$

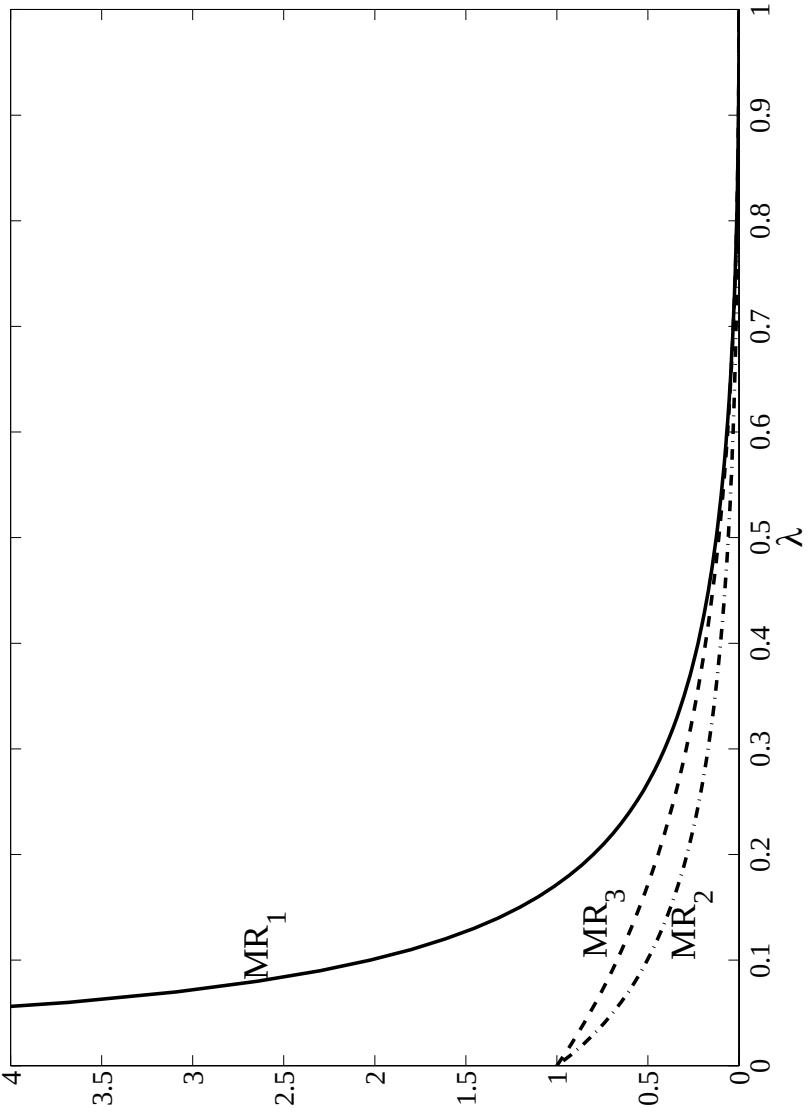


FIGURE 11.7
Plot of the three magnetoresistance ratios MR_1, MR_2 and MR_3 as functions of $\lambda = k_{\downarrow}^F / k_{\uparrow}^F$, where $k_{\downarrow}^F$ and $k_{\uparrow}^F$ are defined in section 11.2.1.

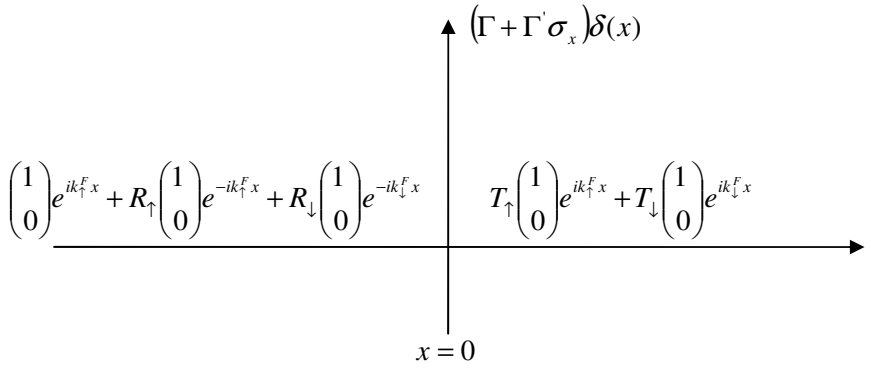

$(\Gamma + \Gamma' \sigma_x)\delta(x)$
$\begin{pmatrix}1\\0\end{pmatrix} e^{ik_\uparrow^F x} + R_\uparrow \begin{pmatrix}1\\0\end{pmatrix} e^{-ik_\uparrow^F x} + R_\downarrow \begin{pmatrix}1\\0\end{pmatrix} e^{-ik_\downarrow^F x}$
$T_\uparrow \begin{pmatrix}1\\0\end{pmatrix} e^{ik_\uparrow^F x} + T_\downarrow \begin{pmatrix}1\\0\end{pmatrix} e^{ik_\downarrow^F x}$
$x = 0$

FIGURE 11.8
The scattering problem for a majority spin incident from the left contact. The magnetizations in the two contacts are assumed to be parallel.

$$T_\uparrow = \frac{\hbar^2 k_\uparrow^F \left(\hbar^2 k_\downarrow^F + i m_0 \Gamma\right)}{A},$$
$$T_\downarrow = \frac{-i\hbar^2 k_\uparrow^F m_0 \Gamma'}{A}, \tag{11.44}$$

where

$$A = \hbar^4 k_\uparrow^F k_\downarrow^F + i\hbar^2 m_0 \Gamma \left(k_\uparrow^F + k_\downarrow^F\right) + m_0^2 \left(\Gamma'^2 - \Gamma^2\right). \tag{11.45}$$

In units of e^2/h, the conductance of the majority spin band is given by

$$G_\uparrow = |T_\uparrow|^2 + \frac{k_\downarrow^F}{k_\uparrow^F} |T_\downarrow|^2. \tag{11.46}$$

It can also be shown that

$$|T_\uparrow|^2 + \frac{k_\downarrow^F}{k_\uparrow^F} |T_\downarrow|^2 + |R_\uparrow|^2 + \frac{k_\downarrow^F}{k_\uparrow^F} |R_\downarrow|^2 = 1, \tag{11.47}$$

which merely expresses the obvious fact that the sum of reflection and transmission probabilities, with and without spin flip, must add up to unity.

The scattering problem described above must be repeated for electrons incident from the minority spin subband. In this case, the reflection and transmission amplitudes are

$$R'_\uparrow = \frac{-i\hbar^2 k_\downarrow^F m_0 \Gamma'}{A},$$

$$R'_{\downarrow} = -1 + \frac{\hbar^4 k_{\uparrow}^F k_{\downarrow}^F + i\hbar^2 m_0 \Gamma k_{\downarrow}^F}{A},$$
$$T'_{\uparrow} = \frac{-i\hbar^2 k_{\downarrow}^F m_0 \Gamma'}{A},$$
$$T'_{\downarrow} = \frac{\hbar^4 k_{\uparrow}^F k_{\downarrow}^F + i\hbar^2 m_0 \Gamma k_{\downarrow}^F}{A}, \tag{11.48}$$

from which the conductance (in units of e^2/h) of the minority spin subband is found to be

$$G_{\downarrow} = \frac{k_{\uparrow}^F}{k_{\downarrow}^F}|T_{\uparrow}|^2 + |T_{\downarrow}|^2. \tag{11.49}$$

Since the majority and minority spins are *orthogonal* in the ferromagnetic contacts, the total conductance (in units of e^2/h) is

$$G_{total} = G_{\uparrow} + G_{\downarrow} = \left(\frac{k_{\uparrow}^F}{k_{\uparrow}^F} + 1\right)|T_{\uparrow}|^2 + \left(\frac{k_{\downarrow}^F}{k_{\downarrow}^F} + 1\right)|T_{\downarrow}|^2. \tag{11.50}$$

Exercise: Using the boundary conditions (11.43), show that the reflection and transmission amplitudes of a majority spin incident from the left contact, as shown in Figure 11.8, are indeed given by equation (11.44). Repeat the exercise for electrons incident from the minority spin subband. This problem can be solved using software tools such as MATLAB or MATHEMATICA.

The total conductance in the parallel configuration G_P can then be calculated by adding the conductances $G_{\uparrow}$ and $G_{\downarrow}$ found above. The scattering problem must then be repeated for the two contacts in the anti-parallel configuration to derive G_{AP} to finally calculate the magnetoresistance ratios MR_1, MR_2 and MR_3.

Numerical examples:
(a) $\Gamma \neq 0$ and $\Gamma' = 0$

When the magnetizations of the two ferromagnetic contacts are parallel and spin-dependent scattering at the interface is neglected, the Landauer conductance of majority and minority spin is equal to 1 (in units of e^2/h) since there is no spin mixing and no interface potential step.

If the magnetization of the right contact is parallel to that of the left contact, show that the transmission coefficient for a majority spin incident from the left contact is given by

$$|T_{\uparrow}|^2 = k_{\uparrow}^{F2} / \left(k_{\uparrow}^{F2} + k_B^2\right), \tag{11.51}$$

where $k_B = m\Gamma/\hbar^2$.

For the minority spin incident from the left contact, we have

$$|T_\downarrow|^2 = k_\downarrow^{F2} / \left(k_\downarrow^{F2} + k_B^2\right). \tag{11.52}$$

The Landauer conductance (in units of e^2/h) for the anti-parallel configuration is therefore given by

$$G_P = |T_\uparrow|^2 + |T_\downarrow|^2 = \frac{k_\uparrow^{F2}}{k_\uparrow^{F2} + k_B^2} + \frac{k_\downarrow^{F2}}{k_\downarrow^{F2} + k_B^2}. \tag{11.53}$$

In the anti-parallel configuration, both $|T_\uparrow|^2$ and $|T_\downarrow|^2$ are given by

$$|T_\uparrow|^2 = \frac{4k_\uparrow^F k_\downarrow^F}{\left[k_\uparrow^F + k_\downarrow^F\right]^2 + 4k_B^2}, \tag{11.54}$$

and therefore the Landauer conductance for the anti-parallel case is given by

$$G_{AP} = \frac{8k_\uparrow^F k_\downarrow^F}{\left[k_\uparrow^F + k_\downarrow^F\right]^2 + 4k_B^2}. \tag{11.55}$$

Using the results above, we plot the three magnetoresistance ratios MR_1, MR_2 and MR_3 in terms of the quantity P defined as

$$P = \frac{k_\uparrow^F - k_\downarrow^F}{k_\uparrow^F + k_\downarrow^F}, \tag{11.56}$$

which characterizes the polarization of the contacts. In Figures (11.9) - (11.11), the different curves in each plot correspond to different values of the ratio $k_B/k_\uparrow^F$, where $k_B = m_0\Gamma/\hbar^2$ characterizes the strength of the spin independent delta scatterer.

The magnetoresistance ratio MR_1 is very small unless P approaches unity. This means MR_1 is tiny unless $k_\downarrow^F$ is close to zero, or the exchange splitting energy Δ is close to the Fermi energy E_F (recall the Stoner-Wohlfarth model of the ferromagnets in Section 11.2.1). For $P = 1$, neither up-spin nor down-spin can be transmitted across the interface and the conductance for the anti-parallel magnetizations is equal to zero, which leads to a divergence of MR_1. This is why MR_1 is sometimes referred to as "inflationary definition of the magnetoresistance" since it is unbounded from above.

In contrast, the magnetoresistance ratios MR_2 and MR_3 have a maximum value of 1 when $P = 1$. As Figures (11.9) - (11.11) indicate, the ratios MR_1, MR_2 and MR_3 are fairly sensitive functions of the strength of the scattering potential representing the barrier between the two ferromagnetic contacts. This trend has been confirmed with much more sophisticated forms of the scattering potential than the one considered here. It has also been found to be

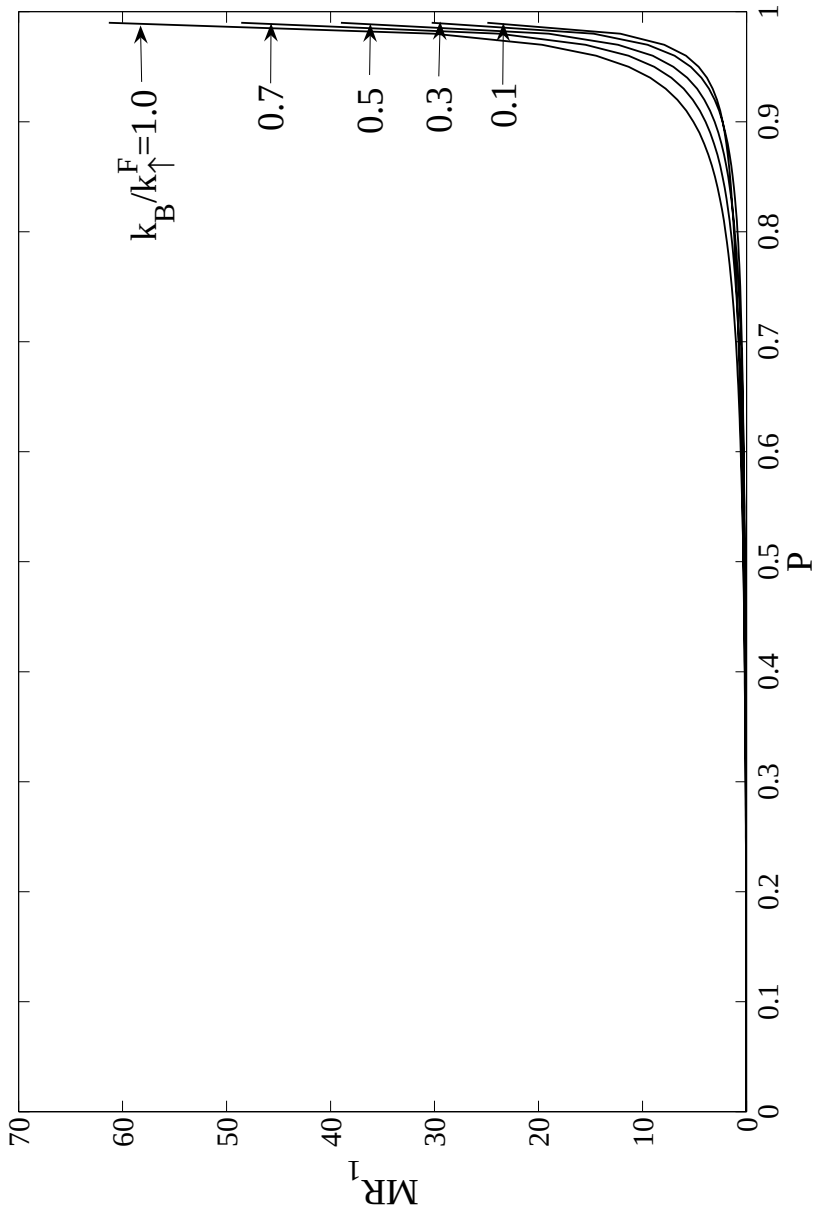


FIGURE 11.9
Plot of the magnetoresistance ratio $MR_1 = (R_{AP} - R_P)/R_P$ as a function of $P = \left(k_\uparrow^F - k_\downarrow^F\right) / \left(k_\uparrow^F + k_\downarrow^F\right)$ for different values of the ratio $k_B/k_\downarrow^F$.

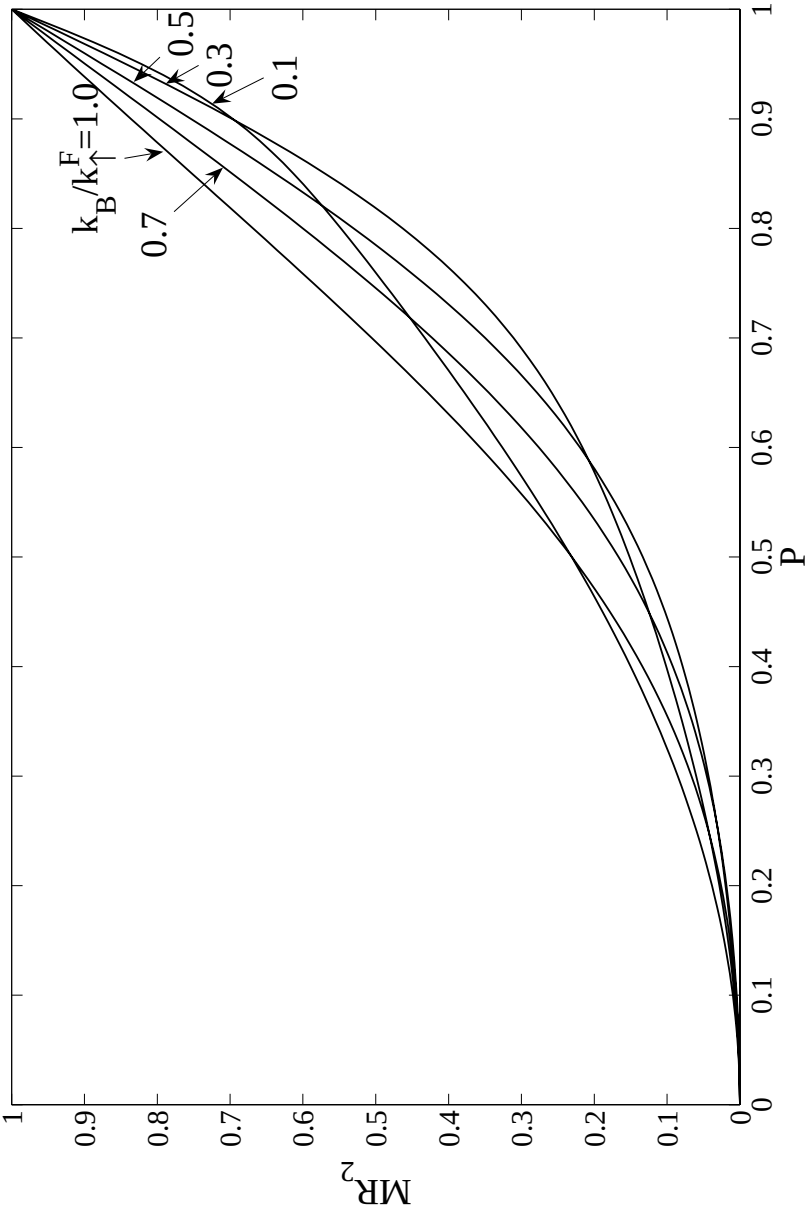


FIGURE 11.10
Plot of the magnetoresistance ratio MR_2 as a function of P.

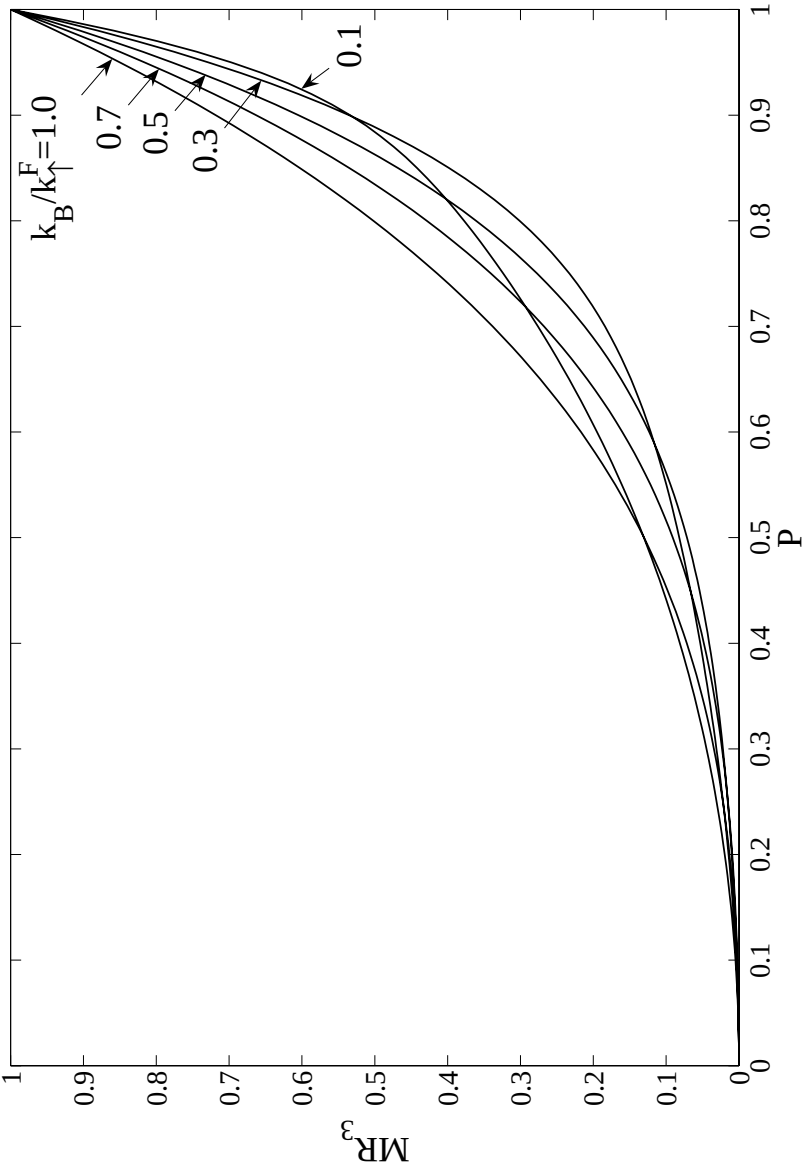


FIGURE 11.11
Plot of the magnetoresistance ratio MR_3 as a function of P.

practically universal according to detailed models developed to treat metallic, semiconducting, insulating and organic materials as the intermediate layer of a spin valve.

(b) $\Gamma \neq 0$ and $\Gamma' \neq 0$

In Figure 11.12, we plot the magnetoresistance ratio MR_3 in Equation (11.20) as a function of the square of the polarization factor P for different values of the ratios $k_B/k_\uparrow^F$ and $k'_B/k_\uparrow^F$, where $k_B = m_0\Gamma/\hbar^2$ and $k'_B = m_0\Gamma'/\hbar^2$ characterize the strength of the spin-independent and spin-dependent scattering potentials at the interface between the two ferromagnetic contacts. Figure 11.12 has a total of nine curves corresponding to values of $k_B/k_\uparrow^F$ and $k'_B/k_\uparrow^F$ equal to $0.1, 0.5$ and 0.9.

Figure 11.12 shows that MR_3 is highly sensitive to the presence of both spin-independent and spin-dependent scattering mechanisms at the interface. Actually, the magnetoresistance MR_3 can become negative which means that $G_P < G_{AP}$ in this case.

There have been reports of both positive and negative magnetoresistances in the literature. These reports can be confusing since different workers adopt different definitions of the magnetoresistance ratios. For instance, if the interaction between two ferromagnetic contacts is such that their magnetization is anti-parallel for a specific barrier, the spin valve will exhibit a lower resistance when parallel magnetization is induced via an external magnetic field. This switching from high to low resistance in the presence of an external magnetic field is referred to as negative magnetoresistance. The results of the simple model studied in this section show that positive magnetoresistance is also possible. Thus, extended measurements over a wide variety of junctions can be very informative about the details of the physical mechanisms involved in spin conduction through the junction.

11.2.4.1 The Jullière formula

In the early years, the magnetoresistance of spin valves was studied systematically by Jullière [3]. In his pioneering work, Jullière prescribed a formula for the junction magnetoresistance ratio (definition MR_2) based on a generalization of the transfer matrix model discussed above. The Jullière formula relates MR_2 to the spin polarizations of carriers at the Fermi level in the ferromagnetic contacts. This formula is often used in the analysis of spin valve experiments and it has been extended to account for spin relaxation in the intermediate paramagnetic layer. Because of its historical importance, we outline the derivation of the Jullière formula here. However, the reader should be warned that the Jullière formula is often used without a careful assessment of its applicability for the specific structure under investigation. In fact, there are some recent reports addressing the limits of validity of the Jullière formula, as increasing numbers of spin valve experiments are being reported.

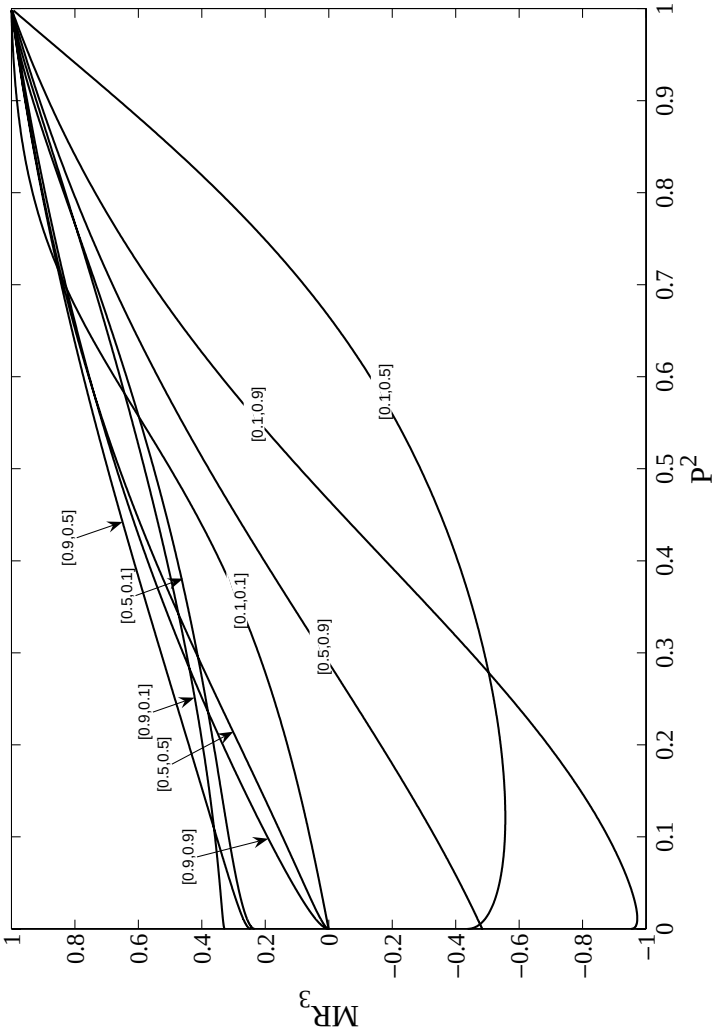


FIGURE 11.12
Plot of the ratio $MR_3 = (G_P - G_{AP})/(G_P + G_{AP})$ versus P^2 for different values of the parameters $\left[k_B/k_\uparrow^F, k'_B/k_\uparrow^F\right]$ characterizing the strength of the spin-independent and spin-dependent scattering potentials between the two ferromagnetic contacts.

In 1975, Jullière experimented with the magnetoresistance of Fe-Ge-Co junctions at $T \leq 4.2K$, as a function of the applied bias between the two ferromagnetic contacts made of two different materials.

To interpret his experimental results, he modified Bardeen's derivation of the tunneling current between two normal paramagnetic metals separated by a thin insulating barrier [15]. Bardeen's formula is very similar to the Tsu-Esaki formula routinely used by device engineers and physicists in the study of resonant tunneling diodes and devices based on resonant tunneling. The basic ingredients are the following: In the absence of an applied voltage, the Fermi levels of the two electrodes flanking a tunnel barrier must be aligned (no current flows). An applied voltage V shifts all the energy levels in one of the electrodes relative to the other by eV. This amounts to an energy shift eV between the two Fermi levels that then causes the tunneling current. In the Bardeen approach, one evaluates the current $I(V, E)$ flowing at a given energy E between the left(L) and right(R) electrodes from Fermi's golden rule, i.e.,

$$I(V, E) \propto |T(E)|^2 D_L(E) D_R(E + eV_{bias})[f(E) - f(E + eV_{bias})], \quad (11.57)$$

where $D_L(E)$, $D_R(E)$ are the densities of states of the left and right electrodes, $|T(E)|^2$ is the square of the tunneling matrix element, and $f(E)$ is the Fermi function.

If the tunneling matrix element is independent of energy over the relevant energy range $\approx eV_{bias}$, then the total tunneling current is obtained by integrating Equation (11.57) with respect to the energy

$$I(V) \propto |T|^2 \int_{-\infty}^{\infty} D_L(E) D_R(E + eV_{bias})[f(E) - f(E + eV_{bias})]dE. \quad (11.58)$$

In the low-bias regime ($V \to 0$) and at low temperatures ($T \to 0$)

$$\lim_{V_{bias}, T \to 0} \frac{[f(E) - f(E + eV_{bias})]}{eV_{bias}} = \delta(E - E_F), \quad (11.59)$$

where E_F is the Fermi energy, so that we find

$$G = \frac{dI}{dV} \propto |T|^2 D_L(E_F) D_R(E_F), \quad (11.60)$$

where $dI/dV = G$ is the zero-bias conductance of the tunneling junction. It is clear from the above equation that in the low-bias regime, $I \propto V$, i.e., the junction displays an ohmic behavior.

Equation (11.60) was applied by Jullière [3] to discuss the magnitude of the observed tunneling magnetoresistance of an FM-I-FM junction with two ferromagnetic (FM) electrodes separated by an insulating (I) barrier. In the case of magnetic electrodes, the densities of states for electrons with spin parallel

$D^{\uparrow}(E)$ and anti-parallel $D^{\downarrow}(E)$ to the magnetization are different because the energy bands of up-spin and down-spin electrons are split by the exchange interaction (recall the Stoner-Wohlfarth bandstructure in Fig. 11.1).

To take into account the dependence of the tunneling current on the relative orientation of the magnetic moments of the two electrodes, Jullière made an additional assumption that electron spin is conserved in tunneling. It then follows that tunneling of up-spin and down-spin electrons are two independent processes, i.e., the tunneling current flows in the up-spin and down-spin channels as if in two wires connected in parallel. Such a two-current model is also used to interpret the closely related giant magnetoresistance effect in magnetic multilayers. Assuming that the magnetic moments of the two electrodes are anti-parallel at zero magnetic field and parallel in a saturating applied field H_s, we find using Equation (11.60) that the conductance of the junction in zero field is

$$G(0) \propto D_L^{\uparrow}(E_F)D_R^{\downarrow}(E_F) + D_L^{\downarrow}(E_F)D_R^{\uparrow}(E_F), \tag{11.61}$$

and its conductance at the saturating field is

$$G(H_s) \propto D_L^{\uparrow}(E_F)D_R^{\uparrow}(E_F) + D_L^{\downarrow}(E_F)D_R^{\downarrow}(E_F). \tag{11.62}$$

As a result, the magnetoresistance ratio is given by

$$MR_2 = JMR = \frac{G(0)^{-1} - G(H)^{-1}}{G(0)^{-1}}. \tag{11.63}$$

Introducing two new parameters P_1 and P_2 to characterize the spin polarization of left and right electrodes,

$$P_1 = [D_L^{\uparrow}(E_F) - D_L^{\downarrow}(E_F)]/[D_L^{\uparrow}(E_F) + D_L^{\downarrow}(E_F)] \tag{11.64}$$

and

$$P_2 = [D_R^{\uparrow}(E_F) - D_R^{\downarrow}(E_F)]/[D_R^{\uparrow}(E_F) + D_R^{\downarrow}(E_F)], \tag{11.65}$$

we find that the magnetoresistance ratio is given by the Jullière formula

$$JMR = \frac{2P_1P_2}{1 + P_1P_2}, \tag{11.66}$$

which we show next.

The parameters P_1 and P_2 can be determined independently from measurements of the tunneling current in FM-I-S junctions, in which one of the electrodes is a ferromagnet (FM) and the other is a superconductor (S).

Substituting the values $G(0)$ and $G(H)$ in Equation (11.63), the junction magnetoresistance ratio is found to be

$$JMR = \frac{\left(\frac{1}{D_L^{\uparrow}(E_F)D_R^{\downarrow}(E_F)+D_L^{\downarrow}(E_F)D_R^{\uparrow}(E_F)}\right) - \left(\frac{1}{D_L^{\uparrow}(E_F)D_R^{\uparrow}(E_F)+D_L^{\downarrow}(E_F)D_R^{\downarrow}(E_F)}\right)}{\frac{1}{\left(D_L^{\uparrow}(E_F)D_R^{\downarrow}(E_F)+D_L^{\downarrow}(E_F)D_R^{\uparrow}(E_F)\right)}}$$

$$= \frac{D_L^{\uparrow}(E_F)[D_R^{\uparrow}(E_F) - D_R^{\downarrow}(E_F)] - D_L^{\downarrow}(E_F)[D_R^{\uparrow}(E_F) - D_R^{\downarrow}(E_F)]}{D_L^{\uparrow}(E_F)D_R^{\uparrow}(E_F) + D_L^{\downarrow}(E_F)D_R^{\downarrow}(E_F)}. \tag{11.67}$$

Dividing numerator and denominator by $[D_L^{\uparrow}(E_F)+D_L^{\downarrow}(E_F)] \times [D_R^{\uparrow}(E_F)+D_R^{\downarrow}(E_F)]$ and rearranging, we find

$$JMR = \frac{P_1 P_2}{\frac{D_L^{\uparrow}(E_F)D_R^{\uparrow}(E_F)+D_L^{\downarrow}(E_F)D_R^{\downarrow}(E_F)}{[D_L^{\uparrow}(E_F)+D_L^{\downarrow}(E_F)].[D_R^{\uparrow}(E_F)+D_R^{\downarrow}(E_F)]}}. \tag{11.68}$$

Finally, multiplying and dividing the last expression by 2 and adding and subtracting 1 in the denominator and simplifying, we get

$$\begin{aligned} JMR &= \frac{2P_1P_2}{\frac{2D_L^{\uparrow}(E_F)D_R^{\uparrow}(E_F)+2D_L^{\downarrow}(E_F)D_R^{\downarrow}(E_F)}{[D_L^{\uparrow}(E_F)+D_L^{\downarrow}(E_F)][D_R^{\uparrow}(E_F)+D_R^{\downarrow}(E_F)]} - 1 + 1} \\ &= \frac{2P_1P_2}{\frac{D_L^{\uparrow}(E_F)D_R^{\uparrow}(E_F)+D_L^{\downarrow}(E_F)D_R^{\downarrow}(E_F)-D_L^{\uparrow}(E_F)D_R^{\downarrow}(E_F)-D_L^{\downarrow}(E_F)D_R^{\uparrow}(E_F)}{[D_L^{\uparrow}(E_F)+D_L^{\downarrow}(E_F)][D_R^{\uparrow}(E_F)+D_R^{\downarrow}(E_F)]} + 1} \\ &= \frac{2P_1P_2}{\frac{D_L^{\uparrow}(E_F)[D_R^{\uparrow}(E_F)-D_R^{\downarrow}(E_F)]-D_L^{\downarrow}(E_F)[D_R^{\uparrow}(E_F)-D_R^{\downarrow}(E_F)]}{[D_L^{\uparrow}(E_F)+D_L^{\downarrow}(E_F)][D_R^{\uparrow}(E_F)+D_R^{\downarrow}(E_F)]} + 1}. \end{aligned} \tag{11.69}$$

Using the definitions of the spin polarization (P_1, P_2) of the contacts, we finally get the compact result

$$JMR = \frac{2P_1P_2}{1 + P_1P_2}, \tag{11.70}$$

which is the famous Jullière formula.

Exercise:

- It is customary to estimate the difference in resistance between the parallel and anti-parallel magnetizations of a tunnel junction (spin valve in which transport takes place through the intermediate layer via tunneling) using the tunneling magnetoresistance ratio (TMR) defined as

$$TMR = \frac{R_A - R_P}{R_P}. \tag{11.71}$$

  Using the Jullière formula, show that

$$TMR = \frac{2P_1P_2}{1 - P_1P_2}. \tag{11.72}$$

- Starting with the Jullière formula, find the expression of the magnetoresistance ratio $(G_P - G_{AP}) / (G_P + G_{AP})$ in terms of P_1 and P_2, the spin polarizations of the ferromagnetic contacts.

11.2.5 Application of the Jullière formula to extract the spin diffusion length in a paramagnet from spin valve experiments

Consider a spin valve consisting of a ferromagnet-paramagnet-ferromagnet structure. Electrons are injected from the left ferromagnetic contact into the paramagnet with a polarization P_1, which is the spin polarization of carriers at the Fermi energy in the left ferromagnet. We assume that there is no loss of spin polarization at the interface. This is actually a pretty good approximation in some cases, e.g., when the paramagnet is a conjugated (organic) polymer that exhibits a "proximity effect", whereby a few atomic layers in contact with the ferromagnet can get spin polarized [16]. The polarization of the injected spins in the paramagnet decay with distance x as $exp[-x/L_s]$ as the spins drift and diffuse (see Chapter 10) across the paramagnetic layer. When the spins arrive at the right ferromagnetic contact, the polarization has decayed to $P_1 e^{-d/L_s}$ where d is the width of the paramagnetic layer. We then assume that there is a tunnel barrier at the right contact through which the spins tunnel. The TMR associated with tunneling through this barrier is given by the Julliére formula (Equation (11.72)) with P_1 replaced by $P_1 e^{-d/L_s}$. Therefore

$$TMR = \frac{2P_1 e^{-d/L_s} P_2}{1 - P_1 e^{-d/L_s} P_2}. \tag{11.73}$$

Thus, from the measured TMR and knowledge of P_1, P_2 and d, we can find the spin diffusion length L_s. This method has been widely used to find the spin diffusion length in many materials.

11.2.6 Spin valve experiments

The typical magnetoresistance trace of a spin valve is shown in Fig. 11.13. One first applies a very high magnetic field on the structure which magnetizes both ferromagnetic contacts in the direction of the field. In this state, the contacts are magnetized parallel to each other, so that the device resistance will be low, assuming that the spin polarizations in the two contacts have the same sign (e.g., cobalt and nickel). The magnetic field is then decreased, swept past zero and made negative (i.e., its direction reversed). At some point, when the coercive field H_{c1} of one of the ferromagnets is reached, this ferromagnet flips magnetization, while the other ferromagnet (which has the higher coercivity H_{c2}) still retains its original direction of magnetization. At this point, the two magnetizations are antiparallel and the resistance increases. Finally, when the magnetic field in the reverse direction increases further, the coercive field of the second ferromagnet is reached and this one flips too. Now, both magnetizations are again parallel and the resistance drops. Therefore, we should see a resistance peak between the coercive fields H_{c1} and H_{c2}. The height of this peak is $\Delta R = R_{AP} - R_P$. If the magnetic field is scanned in

the reverse direction, we will see the same peak between the coercive fields. The background resistance is R_P.

From the measured data, we can find $TMR = \Delta R/R_P$, which then yields the spin diffusion length in the paramagnet from Equation (11.73).

The reader can easily infer that if the two ferromagnetic contacts have opposite signs of spin polarization, then instead of a spin valve "peak", one will observe a spin valve "trough".

11.3 Hysteresis in spin valve magnetoresistance

In real experiments, spin valve effects are sometimes masked by spurious phenomena such as anisotropic magnetoresistance (AMR) that was discovered by Lord Kelvin in 1851, and arises whenever the magnetizations of the ferromagnets are not collinear with the direction of current flow. Hysteresis of the ferromagnets can also complicate matters. These spurious effects are easily eliminated if the combined resistance of the ferromagnets is much smaller than the resistance of the paramagnet. Note that the spin valve device has three resistors in series: the first ferromagnet, the paramagnet and the second ferromagnet. The paramagnet is by far the most resistive if it is a semiconductor or insulator, and the ferromagnets are metals. Therefore, the paramagnet's resistance will dominate. If the resistance peak ΔR is much larger than the total resistance of the ferromagnets, it is most likely not associated with AMR or hysteresis and can be unambiguously ascribed to the "spin valve effect".

In this section, we discuss the hysteresis effects in a ferromagnet since they are unavoidable features of any ferromagnet. A piece of bulk ferromagnet is made up of "domains" that have sizes varying between fractions of a micron to a few microns. Within each domain, the magnetic moments of all the atoms (those that have any residual magnetic moment) points in the same direction, so that each domain has complete spontaneous magnetic ordering and therefore an internal magnetic field. In the "unmagnetized" state, the magnetic moments of different domains point in different directions, so that they vectorially add up to zero and there is no net magnetization of the ferromagnet as a whole. When a sufficiently strong external magnetic field is applied to magnetize the ferromagnet, the magnetic moment of every domain turns in the direction of the field, so that ultimately all domains are magnetized parallel to the field. This is the saturated "magnetized" state.

There are two types of interactions between spins: "exchange" that we have discussed in Chapter 9, and "dipolar" interaction, which we have not discussed. A discussion of dipolar interaction can be found in ref. [2]. It is much weaker, typically a thousand times weaker between neighboring spins in a metal, than exchange. However, the exchange interaction is short-range

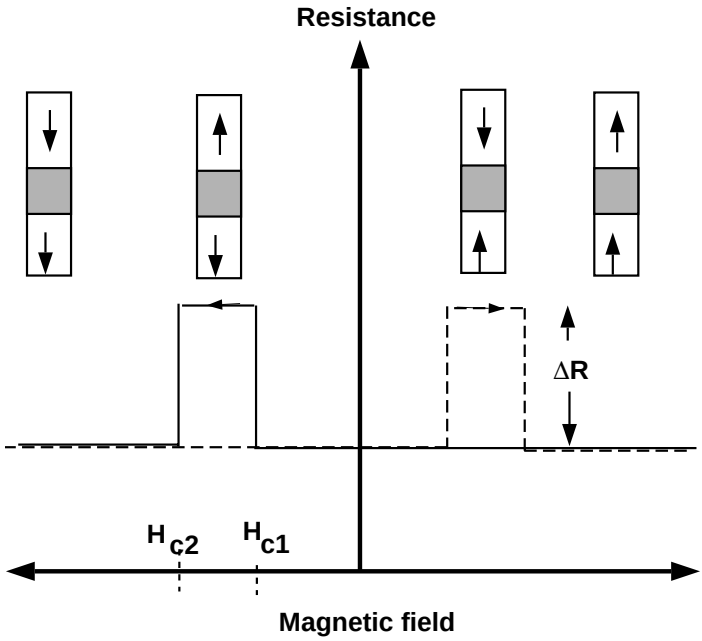


FIGURE 11.13
The ideal magnetoresistance trace measured in a spin valve experiment. A resistance peak is observed between the coercive fields of the two ferromagnets. The spin orientations in the ferromagnetic layers are shown in each region of the magnetic field. We assume that the lower layer has the lower coercivity. The solid trace is for reverse scan of the magnetic field and the broken trace is for the forward scan.

since it depends on the overlap between the wavefunctions of the interacting electrons. It falls off exponentially with increasing separation between the interacting electrons. In contrast, the dipolar interaction is long range and falls off as the inverse cube of the separation. As a result, the dipolar interaction becomes quite important if many spins are involved and may overwhelm the exchange interaction. This fact is responsible for the formation of "domains" in a ferromagnet.

Neighboring domains typically have oppositely directed magnetizations so that electrons bordering a domain wall on two sides have anti-parallel spins. Recall from the Hartree Fock model of Chapter 9 that the exchange energy is non-zero and negative only between parallel spins. Therefore, the spins near the domain boundaries have their total energy raised by the vanishing of the exchange interaction. Spins far from the domain walls are not subjected to this effect since the exchange energy is short range. Therefore, only a small price in exchange energy is paid to form domains. On the other hand, because of the long range nature, the dipolar energy of every spin drops when domains are formed. Therefore, provided the domains are not too small, domain formation is energetically favored. Every spin can lower its (tiny) dipolar energy by forming domains, even though few spins have their (large) exchange energy raised in the process [2].

When a magnetizing field is applied, domains turn in the direction of the field, or domains that are already aligned along the field grow at the expense of unaligned domains. When the magnetizing field is removed, the domains do not necessarily return to their original alignments because crystalline defects may prevent domain wall motion. It then becomes necessary to apply a strong field in the opposite direction to overcome this pinning and restore the original unmagnetized state. This phenomenon is known as *hysteresis* and the magnetic field required to reduce the magnetic moment to zero (corresponding to the original unmagnetized state) is known as the *coercive field*. In the magnetic recording industry, there is usually a great deal of desire to increase the coercive field since then small magnetic perturbations due to noise could not flip stored data and corrupt the memory. However, too large a coercive field is also not a good thing since that requires a strong magnetic field (and hence significant power consumption) to erase old data or write new data.

Hysteresis is usually manifested in magnetic moment versus magnetic field (M-H) measurements. A typical curve for cylindrical iron nanodots of 10 nm diameter and length varying from 200-600 nm is shown in Fig. 11.14. The value of the coercive field, the saturation magnetization and the shape of the hysteresis loop ("squareness") are determined by the precise domain mechanics and magnetic reversal mechanisms.

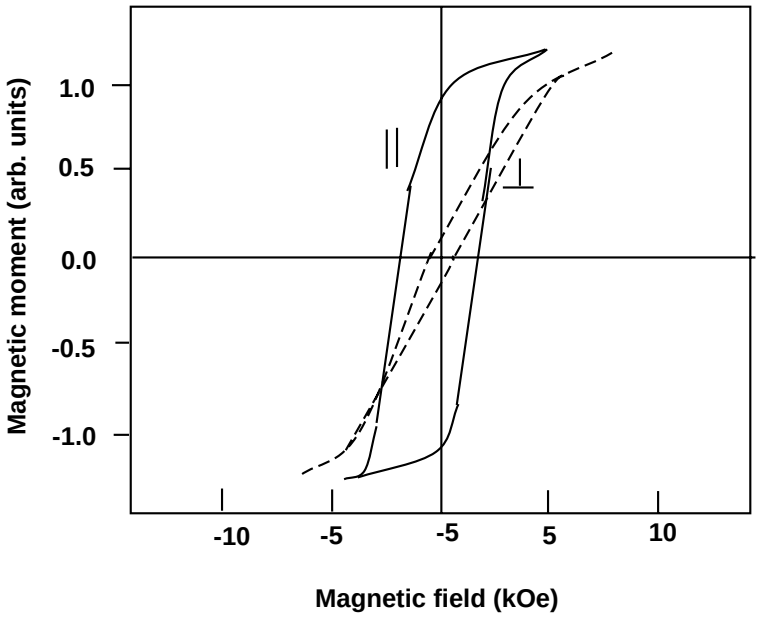


FIGURE 11.14
The M-H loop for iron nano-cylinders of diameter 10-nm and length varying between 100 and 250 nm, showing hysteresis. The solid lines denote the loop obtained when the applied magnetic field is along the axis of the cylinder and the broken lines denote the loop obtained when the magnetic field is transverse to the axis.

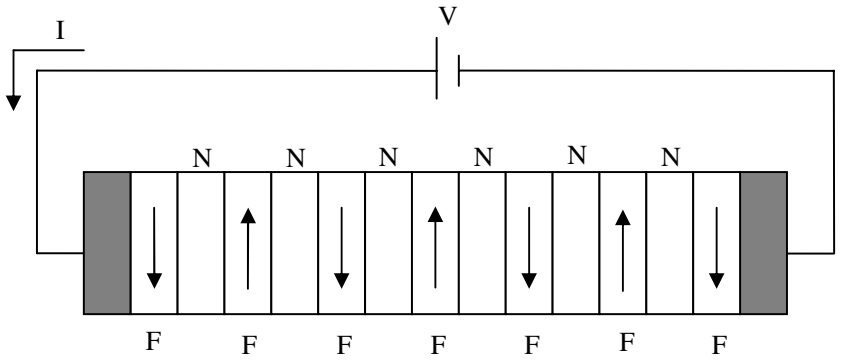


FIGURE 11.15
Spin-valve-superlattice shown in the anti-parallel configuration.

11.4 Giant magnetoresistance

Giant magnetoresistance devices are an extension of spin valves. One might wonder if any benefit would accrue from increasing the number of layers in the basic unit cell of a spin valve, while keeping the width of each layer a few nanometers thick. This would result in forming a "spin-valve-superlattice", as depicted in Fig. 11.15.

Let us use the simple two-resistor model discussed earlier to calculate the resistance of the finite superlattice composed of N different layers.

If N is odd, i.e., the end layers are ferromagnetic, it is easy to show that, in the parallel configuration,

$$R_P = \frac{NR_\uparrow NR_\downarrow}{NR_\uparrow + NR_\downarrow} = N\frac{R_\uparrow R_\downarrow}{R_\uparrow + R_\downarrow}. \tag{11.74}$$

In the anti-parallel configuration, i.e., when the magnetization directions of the magnetic layers in adjacent ferromagnetic layers are anti-parallel

$$R_{AP} = \frac{\left[\frac{N-1}{2}R_\uparrow + \frac{N+1}{2}R_\downarrow\right]\left[\frac{N-1}{2}R_\downarrow + \frac{N+1}{2}R_\uparrow\right]}{N\left(R_\uparrow + R_\downarrow\right)}, \tag{11.75}$$

and if N is large, then

$$R_{AP} = \frac{N}{4}\left(R_\uparrow + R_\downarrow\right). \tag{11.76}$$

It is not difficult to prove that similar expressions hold if N is even and large.

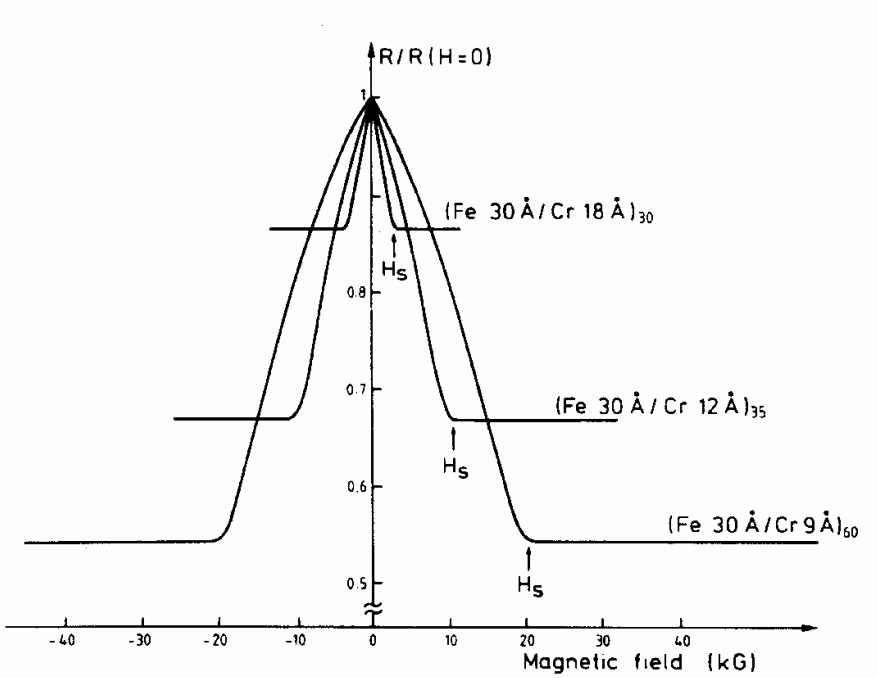


FIGURE 11.16
Magnetoresistance of three Fe/Cr superlattice units at a temperature of 4.2 K. The current and applied magnetic field are along the same [110] axis in the plane of the layers (CIP geometry). Reprinted with permission from M. N. Baibich, et al., Phys. Rev. Lett., **61**, 2472 (1988). Copyright (1988) American Physical Society. http://link.aps.org/abstract/PRL/v61/p2472.

Using the above results for R_P and R_{AP}, it is easy to show that the values of the magnetoresistance ratios MR_1, MR_2 and MR_3 are identical to those derived for the basic spin valve composed of just three layers.

Based on the simple two resistor model, it would appear that there is no additional advantage in building a spin-valve-superlattice and studying its magnetoresistance properties. However, we know that the two-resistor model is oversimplified. Therefore, its predictions are dubious.

In the late 1980s, two groups performed magnetoresistance measurements on multilayered spin-valve superlattices and came across a rather surprising result [4, 5]. Some of the experimental data from ref. [4] are shown in Fig. 11.16.

The resistance of magnetic metallic multilayers dropped to less than one-half of their original value in a weak magnetic field. This huge negative magnetoresistance (MR) of multilayers is called *giant magnetoresistance* (GMR) because it is much larger than the intrinsic MR of the constituent bulk materials. The discovery of the GMR effect in Fe/Cr magnetic multilayers in 1988 [4, 5] spawned a great deal of research on layered magnetic materials systems. Its importance has now been recognized by the Nobel Prize in Physics awarded to authors from the two groups in refs. [4, 5].

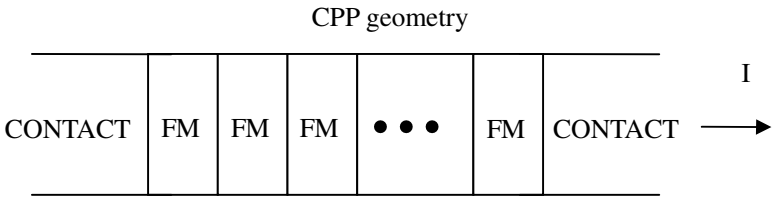
The basic explanation for the GMR phenomenon is qualitatively similar to the one given earlier for the operation of a spin valve composed of just three layers. It is a consequence of the non-equivalence of the two *spin-channels* and the resulting change in the magnetic state of the sample in the presence of a small external magnetic field. The presence of a small external magnetic field forces the magnetizations of the FM-layers to be oriented in parallel and consequently the resistance of the superlattice can drop to less than one-half its value at zero external magnetic field.

Most of the GMR experiments were originally carried out with the current flowing in the plane of the multi-layers, a geometry referred to as current-in-plane (CIP) geometry, as illustrated in Fig.11.17(b). This configuration results in fairly large values of the sample resistance since the sample length (between the current injecting and detecting contacts) is typically orders of magnitude larger than the layer thicknesses. On the other hand, in the so-called current-perpendicular-to-plane (CPP) geometry illustrated in Fig.11.17(a), the length of the sample is just the sum total of the layer thicknesses which is typically much smaller than the film's lateral dimensions. Hence, this configuration results in ultra-low sample resistances, which can only be measured with extremely sensitive equipment. Therefore, the CIP configuration is preferred over the CPP configuration, even though the latter may result in larger relative changes in the resistance in a magnetic field.

Resistor Network Theory of GMR

For a sophisticated treatment of GMR, which is capable of explaining the experimentally observed strong dependence of magnetoresistance on the number

(a)



(b)

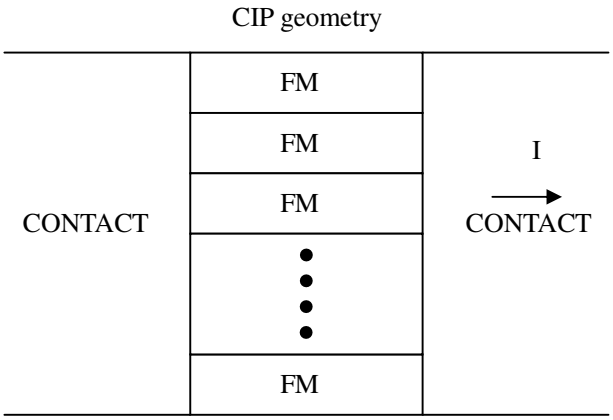


FIGURE 11.17
Magnetoresistance measurements are either performed in the (a) current perpendicular to the plane (CPP) or in the (b) current parallel to the plane (CIP) geometry. Reproduced with permission fromJ. Mathon, "Phenomenological Theory of Giant Magnetoresistance" in ***Spin Electronics***, Eds. M. Ziese and M. J. Thornton, (Springer Verlag, New York, 2001). Copyright(2001) Springer Verlag.

of layers in a GMR sample, we need to incorporate the effect of spin-dependent scattering to determine the resistance of a magnetic multilayer in its "ferromagnetic" (neighboring layers parallel) and "antiferromagnetic" (neighboring layers anti-parallel and alternating layers parallel) configurations. The reader is referred to Chapter 4 of ref. [18] where Mathon has developed an extension of the resistor network theory of GMR which includes the effect of bulk and interface scattering for both the CPP and CIP configurations based on the two fluid model described earlier. Increasingly more sophisticated models of GMR have been developed over the years which are fairly good at explaining most experimental data [17, 19, 20].

11.4.1 Applications of the spin valve and GMR effects

The demonstration of the GMR effect has been one of the early successes of spintronics. It has made a rapid transition from discovery to commercialization [21, 22]. The first commercial product based on GMR was a magnetic field sensor released in 1994 [23], followed shortly in 1997 by the production by IBM of the first GMR read head for reading data stored in magnetic hard disks [24]. In the 1990s, GMR-based read heads replaced earlier read heads based on the anisotropic magnetoresistance effect[§]. Today, GMR based read heads are frequently found in laptop or desktop computers, and music/video storage devices such as Apple iPods. In the 21st century, TMR based read heads began to displace GMR based read heads. The typical TMR device is essentially a thin insulating layer (of thickness about 1 nm) sandwiched between two ferromagnetic contacts. At the time of writing this textbook, the record magnetoresistance exhibited by a TMR device is an astounding 472% at room temperature [25].

The first GMR based random access memory (RAM) chips were produced by Honeywell in early 1997. MRAM chips based on TMR devices are now marketed by several companies, such as Freescale Semiconductor in the US. The impact of GMR/TMR technology on the sensor market is currently estimated worldwide at several $100 million a year [26].

11.4.1.1 Magnetic read heads and magnetic random access memories

Practical spin valves are typically constructed such that the magnetic moment in one of the ferromagnetic layers is more easily reversed than in the other, by an external magnetic field. These materials are referred to as *soft* and *hard* magnetic layers, respectively. The soft layer can be made very sensitive to the presence of a small external magnetic field and acts as the control terminal to allow or block the current flow through the spin valve.

[§]The anisotropic magnetoresistance effect typically produces a magnetoresistance of about 2%.

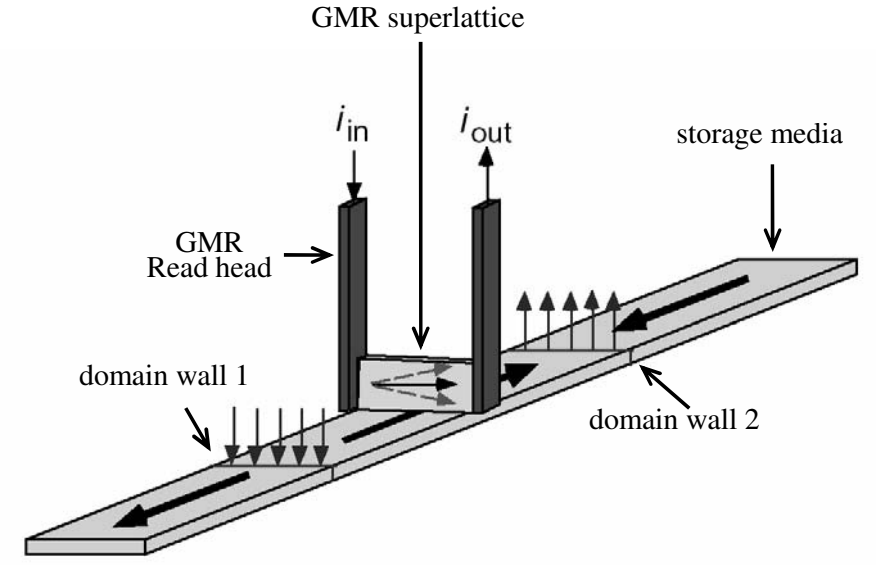


FIGURE 11.18
Schematic of a GMR read head composed of a ferromagnetic/normal-metal superlattice in which the adjacent ferromagnetic layers consist of hard and soft magnetic materials. The magnetization direction of the soft layers is slightly shifted either up or down in response to the magnetic field lines emanating from the domain walls existing at the interface between the magnetic domains in the storage media. The resulting small change in the resistance of the GMR read head is sensed as a small increase or decrease in the current flowing through the GMR element. Reproduced from ref. [24] with permission from the American Association for the Advancement of Science (copyright American Association for the Advancement of Science).

A schematic representation of one of the first magnetic read heads based on the GMR phenomenon is shown in Figure 11.18. The goal of the read head is to sense the magnetic bits that are stored in the storage media (tape or disk). The bits consist of a sequence of magnetized regions or domains whose magnetization directions correspond to either bit 0 or bit 1. The two magnetization directions corresponding to the two bits are anti-parallel. The magnetic field lines emanating from the magnetic domains are only substantial in the vicinity of the domain walls separating two adjacent domains. When the heads (north poles) of two adjacent domains meet (domain wall 2 in Fig. 11.18), uncompensated positive poles in the domains generate magnetic field lines pointing upwards on the surface passing under the GMR read head. In contrast, the magnetic field lines point downward (near domain wall 1 in Fig. 11.18) where the tails (south poles) of adjacent domains meet.

The GMR sensing element is fabricated in such a way that the magnetic moment in the soft ferromagnetic layers is parallel to the plane of the storage media in the absence of any external magnetic field. The magnetic moments in the hard ferromagnetic layers are oriented perpendicular to the plane of the media. When the GMR read head passes over domain wall 1, the field lines in its vicinity pushes the magnetization of the soft layers down, whereas it is pushed up when passing over domain wall 2. As a result, there is a change in the resistance of the GMR element. An increase is associated with more anti-aligned ferromagnetic layers, while a decrease is associated with more aligned layers. The best design of such GMR read heads is one for which a maximum rate of change in resistance can be observed for a small change in the sense field. In state-of-the-art GMR read heads, changes in resistance of about 1% per Oersted are readily achieved.

In 1997, Honeywell demonstrated the first *non-volatile* MRAM based on GMR elements fabricated with standard lithography. These devices exhibited access (read/write) speed and density comparable to semiconductor memory at that time. The basic structure is shown in Fig. 11.19. It consists of GMR elements placed in series and separated by lithographically delineated regions to form a *sense line*. The sense line is the information storing media (tape or disk) with a resistance that is the sum of the resistance of its GMR elements. When active, a current is passed through the sense line and amplifiers at the ends of the line detect the change in resistance in the GMR elements. The resistance of each individual element can be changed through an array of overlay lines (*write* lines) and underlay lines (*bit* lines) to address each individual GMR component. The networks of lines are electrically insulated, but when current pulses are run through them, they generate magnetic fields that can act on the GMR elements. A typical addressing scheme uses "half-select" pulses in the overlay and underlay lines, i.e., the magnetic field produced by a word-line pulse is one-half of that needed to reverse the magnetization in the soft layers of each GMR element. Hence, wherever two lines (word and bit) in the grid overlap, the two half-select pulses they produce generate a combined magnetic field sufficient to selectively reverse a soft layer (or, at

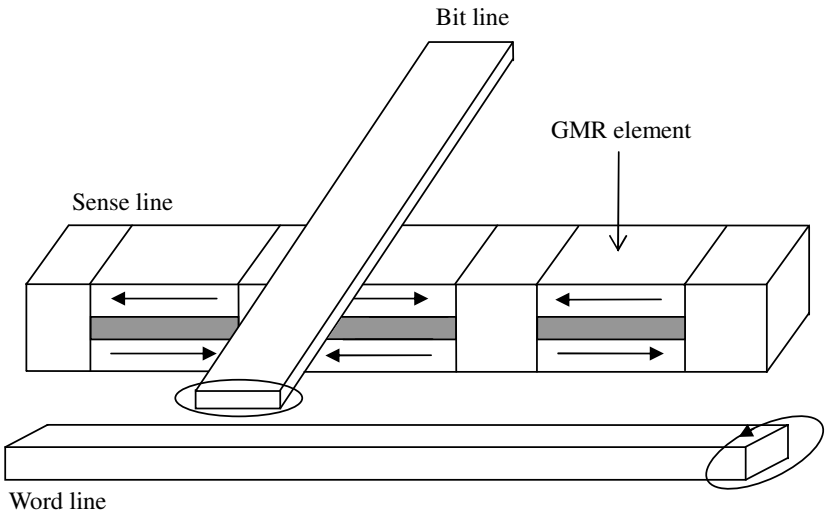


FIGURE 11.19
Schematic illustration of a magnetic random access memory (MRAM) consisting of GMR elements connected in series. The writing and reading of each element is performed via the combination of applied magnetic fields generated by currents driven through the ***Word*** and ***Bit*** lines below and above the GMR elements. Reproduced from ref. [24] with permission from the American Association for the Advancement of Science (copyright American Association for the Advancement of Science).

higher current levels, sufficient to reverse a hard layer also). Typically, a half-select pulse rotates the magnetization in the soft magnetic layer by 90° and the half-select pulse in the other line completes the magnetization rotation to 180° compared to its original state. Through this grid of word and bit lines, any GMR element in the array can be addressed either to store information (write data) or to read its magnetic state (read data) [24].

11.5 Spin accumulation

In this Chapter, we had introduced the concept of spin injection from a ferromagnet into a paramagnet. Spin injection will accumulate spin in the paramagnet near the interface, and the accumulated spin will gradually decay with distance away from the interface. We will now examine this decay and extract a "spin accumulation length" which is the characteristic distance over which the accumulated spin decays to 1/e times its magnitude at the interface.

Let us consider a bulk metal or semiconductor in which the energy dispersion relation is assumed to be parabolic for the sake of simplicity. If the material is non-magnetic, the two spin channels have the same mobility and, if we focus on the (k_x, k_y) plane, the Fermi energy surface for both spin channels will be shifted as shown in the top part of Fig. 11.20 in the presence of a small external electric field. The shift Δk in momentum space of the up-spin and down-spin Fermi surfaces must be such that the following equation is satisfied

$$F = -eE = \hbar \frac{dk}{dt} = \hbar \frac{\Delta k}{\tau_m}, \tag{11.77}$$

where F is the force on the carrier, E is the external electric field, e is the magnitude of the electronic charge, τ_m is the electron scattering time which is related to the mobility of the carrier by $\mu = e\tau_m/m^*$, and m^* is the electron effective mass.

In a ferromagnetic metal, the two spin channels will have different mobilities (recall Mott's two fluid model of spin transport discussed earlier) and the shift of the Fermi surface associated with each spin channel will be different, as illustrated in the bottom part of Fig. 11.20.

Let us consider a ferromagnet/paramagnet heterostructure. We will model the ferromagnet based on the Stoner-Wohlfarth model. We will assume that the up-spin electrons are the majority in the ferromagnet so that the density of up-spin electrons in the ferromagnet is much larger than that of the down-spin electrons. Consequently, the up-spin electrons will be the major contributors to any current injected by the ferromagnet. As a result, there will be a surplus of up-spin electrons in the paramagnet near the interface with the ferromagnet. This will cause *spin accumulation* in the paramagnet near the interface and an

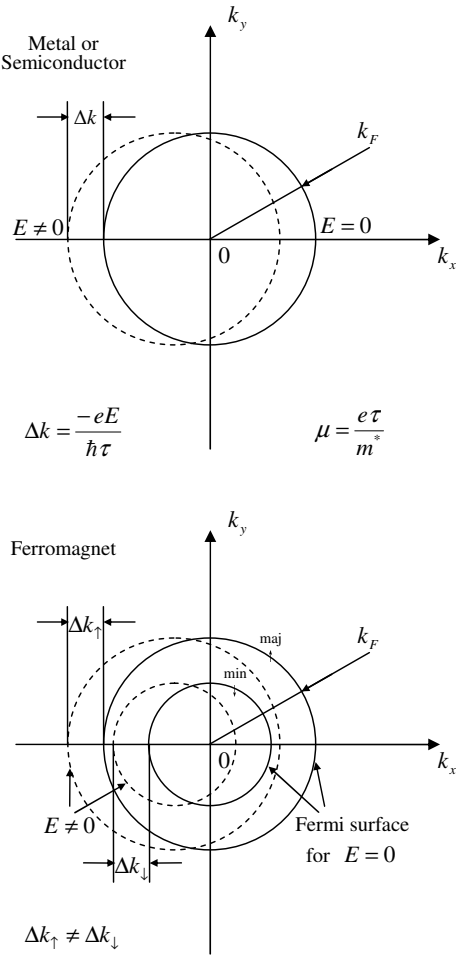


FIGURE 11.20
(Top) Shift of the Fermi surface in a paramagnetic metal or semiconductor in the presence of a constant external electric field leading to current conduction. (Bottom) Generalization to the two-fluid (or Mott model) of the ferromagnet in which the up-spin and down-spin electrons have different mobilities.

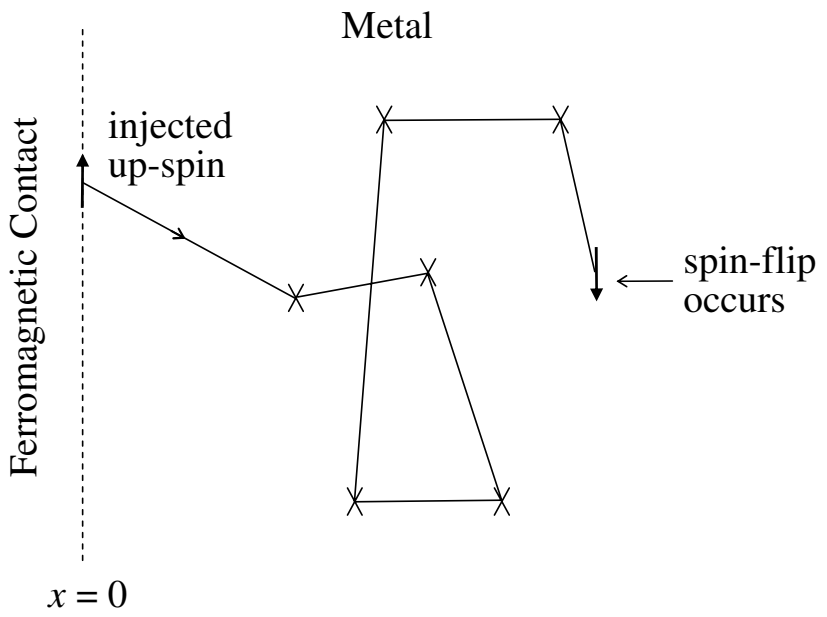


FIGURE 11.21
Illustration of the spin injection mechanism for a up-spin electron injected from a ferromagnetic contact into a paramagnet. After a random walk in the paramagnet where it suffers N momentum relaxing scattering events (indicated by crosses), the spin of the injected electron is eventually flipped. The average (over an ensemble of many injected carriers) distance from the interface for spin flip to occur is roughly the ***spin diffusion length***.

associated magnetic moment per unit volume. Under steady-state condition, the spin accumulation cannot extend into the paramagnet indefinitely since some of the up-spin electrons will eventually be converted into down-spin electrons as a result of spin-flip scattering events. In fact, far into the bulk of the paramagnet, the population of up-spin and down-spin electrons should be the same; therefore, we expect the spin accumulation to decay with distance as we move away from the interface. Hereafter, we give a simple argument to derive an expression for the length scale over which the spin accumulation decays away from the interface.

Based on the drift-diffusion model discussed in Chapter 10, we expect the spin accumulation to decay exponentially away from the interface as $\sim \ e^{-x/\lambda_{sd}}$ where λ_{sd} is the *spin accumulation length*. The latter can be estimated using the following simple method.

As illustrated in Fig. 11.21, an up-spin electron crossing from the ferromagnet into the non-magnetic material will undergo N momentum-changing

collisions before being flipped. We label this average spin flip time $\tau_{\uparrow\downarrow}$. By definition, the average distance between momentum scattering collisions is λ_f, the mean free path. If the injected electron is allowed to move at random equally in all three dimensions after each collision, then the average distance which the spin penetrates into the non-magnetic material (perpendicular to the interface) or spin accumulation length λ_{sd} is given by

$$\lambda_{sd} = \lambda_f \sqrt{N/3}. \tag{11.78}$$

On the other hand, the total distance traveled by the injected up-spin electron is $N\lambda_f$ which in turn equals the injected electron's velocity at the Fermi level (Fermi velocity, v_F) times the spin-flip time τ_s ($=\tau_{\uparrow\downarrow} = \tau_{\downarrow\uparrow}$)[¶], i.e.,

$$N\lambda_f = v_F \tau_s. \tag{11.79}$$

Eliminating the number N of collisions between the last two equations yields the expression

$$\lambda_{sd} = \sqrt{\frac{\lambda_f v_F \tau_s}{3}}. \tag{11.80}$$

This last relation clearly shows that λ_{sd} and λ_f are related but not equal. For example, if we dope silver with increasing levels of a non-magnetic impurity (such as gold), λ_{sd} will drop not only because of a decrease in the electron mean free path λ_f with increasing impurity concentration, but also because a larger concentration of impurities will lead to a reduced spin-flip time τ_s as a result of increased spin-orbit scattering due to the heavy gold atoms [27].

Warning: The reader who has already perused and understood the material in Chapter 8 will be immediately bothered by the preceding discussion. It appears that

$$\frac{\tau_s}{\tau_m} = N, \tag{11.81}$$

where τ_m is the momentum relaxation time. The last equation tells us that the spin relaxation time is directly proportional to the momentum relaxation time. This may be true for the Elliott-Yafet mechanism, but *is certainly not true* for the D'yakonov-Perel' mechanism. Spin relaxation due to the Elliott-Yafet mechanism is, in fact, associated with momentum relaxation and it is known that the spin relaxation time and momentum relaxation time are proportional to each other (recall Chapter 8)). On the other hand, the spin relaxation time associated with D'yakonov-Perel' relaxation is inversely proportional to momentum relaxation time, which immediately contradicts Equation (11.81). Therefore, the picture presented here has no applicability to D'yakonov-Perel'

[¶]We are assuming elastic collisions here, so that the magnitude of the electron's velocity is invariant. We are also assuming that the velocity with which the carriers are injected is the Fermi velocity v_F.

relaxation. In fact, as we have seen in Chapter 10, the entire drift-diffusion model of spin transport is practically invalid if the primary spin relaxation mechanism is D'yakonov-Perel'. The reader should bear this in mind whenever the drift-diffusion model is encountered.

From the previous analysis, we can estimate the importance of spin accumulation in a metal for typical current densities across a ferromagnet/paramagnet interface. Defining the spin polarization α of the injected current as

$$\alpha(x) = \left(J_\uparrow(x) - J_\downarrow(x)\right)/J, \tag{11.82}$$

where J is the total current density $J_\uparrow(x) + J_\downarrow(x)$ across the interface with x being the direction of the normal to the interface. Since J does not depend on x, the net spin accumulation in the metal for a current J flowing through the junction can be found by equating the net spin injection across the interface

$$\left[\frac{d}{dt}\left(n_\uparrow - n_\downarrow\right)\right]_{x=0} = \frac{A\alpha(0)J}{e} \tag{11.83}$$

(where A is the cross-sectional area of the junction) to the decay rate of the overall spin concentration in the entire volume in the paramagnet where spin accumulation occurs. Assuming an exponential decay of the spin accumulation with characteristic decay constant λ_{sd}, we get

$$\left[\frac{d}{dt}\left(n_\uparrow - n_\downarrow\right)\right]_{x=0} = \frac{A}{\tau_s}\int_0^\infty \left(n_\downarrow - n_\uparrow\right) dx. \tag{11.84}$$

Since

$$n_\uparrow(x) - n_\downarrow(x) = n_0 e^{-x/\lambda_{sd}}, \tag{11.85}$$

where $n_0 = n_\uparrow(0) - n_\downarrow(0)$, we get

$$n_0 = \frac{\alpha(0)J\tau_s}{e\lambda_{sd}} = \frac{3\alpha(0)J\lambda_{sd}}{ev_F\lambda_f}. \tag{11.86}$$

.

For $\alpha(0) = 1$ (half-metallic contact), $v_F = 10^6 m/s$, $\lambda_f = 5\ nm$, $\lambda_{sd} = 100\ nm$, and a typical current density $J = 10^3 A/cm^2$, we get $n_0 = 4 \times 10^{22} m^{-3}$. Since in normal metals, the electron concentration is of the order of several $10^{28} m^{-3}$, the net spin accumulation in a normal metal (paramagnet) is typically small with only one part in 10^6 of the electrons being spin polarized. The magnetic field B associated with this spin accumulation is given by

$$B = \mu_0 M = \mu_0 n_0 \mu_B \simeq 10^{-9}\ Tesla, \tag{11.87}$$

which is very small compared to the magnetic field due to the current flowing through the interface and generating the spin accumulation.

11.6 Spin injection across a ferromagnet/metal interface

In 1987, Van Son et al. [28] first pointed out that, because of spin accumulation, whenever a current flows from a ferromagnet into a paramagnet, the distribution of up-spin and down-spin carriers has to change. A similar phenomenon occurs across a superconductor-normal metal interface, where a conversion between normal current and supercurrent is accompanied by an electrochemical potential difference between quasiparticles and Cooper pairs near the interface.

For simplicity, we consider a ferromagnet/normal metal interface in the plane x=0 and assume that the system is homogeneous in the y- and z-directions. The ferromagnet fills the half-space ($x < 0$) and the normal metal is located in the half-space ($x > 0$), as shown in Fig. 11.22. The thickness of each material is assumed to be greater than their respective spin diffusion length and a current of electron is assumed to flow from the left to right in the x-direction only, i.e., the positive terminal of the voltage source (battery) is connected to the normal metal and the negative terminal to the ferromagnetic contact.

If in both contacts, the rate of scattering events that do *not* flip spin far exceeds the spin-flip rates at any coordinate point x, we can define individual electrochemical potential for the up-spin ($\mu_\uparrow$) and down-spin ($\mu_\downarrow$) channels. These electrochemical potentials are expected to be spatially varying and quite different from each other close to the interface. To quantify the difference between them, the conductivity and current density are separated into two components by introducing the variables $\alpha(x)$ and $\beta(x)$, such that

$$\sigma_\uparrow(x) = \alpha(x)\sigma, \tag{11.88}$$

$$\sigma_\downarrow(x) = (1 - \alpha(x))\sigma, \tag{11.89}$$

and

$$J_\uparrow(x) = \beta(x)J, \tag{11.90}$$

$$J_\downarrow(x) = (1 - \beta(x))J, \tag{11.91}$$

where J is the total current density flowing across the interface. In steady state, J is independent of x as a result of current continuity.

At a distance far from the interface (exceeding the spin accumulation length in either material), the two electrochemical potentials $\mu_\uparrow(x)$ and $\mu_\downarrow(x)$ will converge towards each other since the up-spin and down-spin populations should be near their equilibrium values, i.e.,

$$\beta(-\infty) = \alpha_F, \tag{11.92}$$

and

$$\beta(+\infty) = \alpha_N = 0.5, \tag{11.93}$$

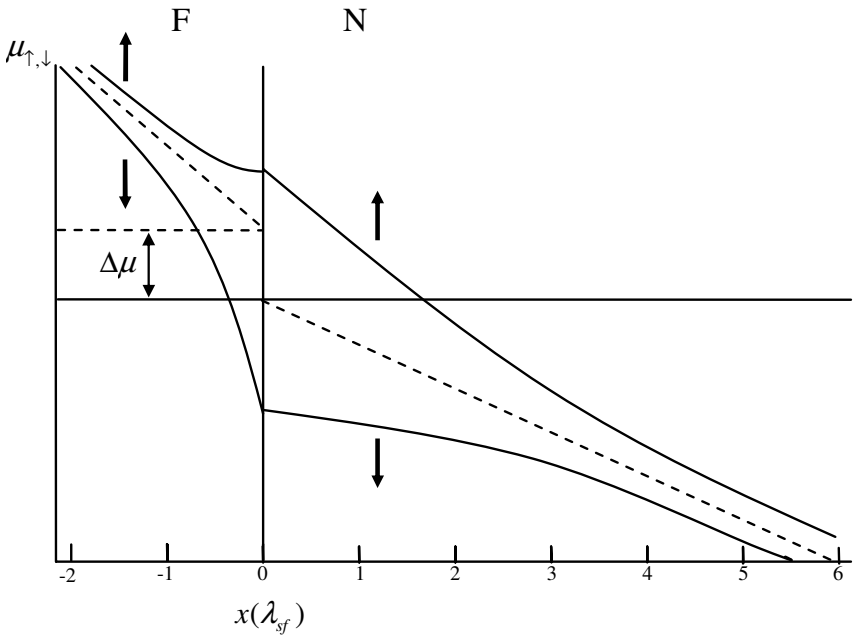


FIGURE 11.22
Illustration of the electrochemical potential difference for up-spin and down-spin electrons near a ferromagnetic metal/non-magnetic metal interface when a current flows through the junction. The tilted dashed line represents the weighted electrochemical potential, $\mu_0 = \alpha(x)\mu_\uparrow(x) + (1-\alpha)\mu_\downarrow(x)$. Reproduced with permission fromH. X. Tang, et al., "Spin Injection and Transport in Micro- and Nano-scale Devices" in ***Semiconductor Spintronics and Quantum Computation***, Eds. D. D. Awschalom, N. Samarth and D. Loss (Springer-Verlag, Berlin, 2002). Copyright(2002) Springer-Verlag.

where the subscripts 'F' and 'N' stand for ferromagnet and normal metal (or non-magnet), respectively.

Furthermore, $\alpha(x)$ is expected to change abruptly at the interface due to the difference in conductivities of the two materials but $\beta(x)$ must be continuous in the absence of spin-flip scattering mechanisms at $x = 0$. Consequently, in regions that are a few diffusion lengths away from the interface, $d\beta/dx \neq 0$ and $\mu_\uparrow(x) \neq \mu_\downarrow(x)$. This electrochemical potential difference is the driving force for the spin current conversion across the interface.

At steady state, the electrochemical potential difference obeys the diffusion equation

$$\frac{\mu_\uparrow - \mu_\downarrow}{\tau_s} = D\frac{d^2(\mu_\uparrow - \mu_\downarrow)}{dx^2}, \tag{11.94}$$

where τ_s is the spin-flip scattering rate and D is the *charge* diffusion coefficient which, in the non-magnetic metal, is given by

$$D = \frac{1}{3}v_F\lambda_f. \tag{11.95}$$

In the ferromagnet, the charge diffusion coefficient D_F is estimated using a weighted average

$$D_F = (1 - \alpha_F)D_{F,\uparrow} + \alpha_F D_{F,\downarrow}, \tag{11.96}$$

where $D_{F,\uparrow}$ and $D_{F,\downarrow}$ are the charge diffusion coefficients for up-spin and down-spin electrons that are not necessarily equal to each other. A derivation of Equation (11.94) is given in a series of problems at the end of the chapter.

When a small bias is applied between the contacts to the ferromagnet and normal metal, a general solution to the diffusion equation (11.94) can be found by assuming the following spatial dependence of the electrochemical potentials:

In the ferromagnetic contact, the electrochemical potentials for the up-spin and down-spin channels are of the form

$$\begin{aligned}\mu_{F\uparrow}(x) &= \mu_1 + a_1 x + c_\uparrow^F e^{x/\lambda_{sd,F}},\\ \mu_{F\downarrow}(x) &= \mu_1 + a_1 x + c_\downarrow^F e^{x/\lambda_{sd,F}},\end{aligned} \tag{11.97}$$

while in the non-magnetic metal, they have the form

$$\begin{aligned}\mu_{N,\uparrow}(x) &= \mu_2 + a_2 x + c_\uparrow^N e^{-x/\lambda_{sd,N}},\\ \mu_{N,\downarrow}(x) &= \mu_2 + a_2 x + c_\downarrow^N e^{-x/\lambda_{sd,N}},\end{aligned} \tag{11.98}$$

where, again, the subscripts 'F' and 'N' refer to ferromagnet and non-magnet, respectively.

The spin current densities in both spin channels are related to the electrochemical potentials according to the relations [29]:

$$J_{F,\uparrow}(x) = -\frac{\beta_F}{e}\sigma_F\frac{d}{dx}\mu_{F,\uparrow}(x), \tag{11.99}$$

and

$$J_{F,\downarrow}(x) = -\frac{1-\beta_F}{e}\sigma_F \frac{d}{dx}\mu_{F,\downarrow}(x), \tag{11.100}$$

in the ferromagnetic contact, while in the normal metal, these relations are [29]

$$J_{N,\uparrow}(x) = -\frac{\sigma_N}{2e}\frac{d}{dx}\mu_{N,\uparrow}(x), \tag{11.101}$$

and

$$J_{N,\downarrow}(x) = -\frac{\sigma_N}{2e}\frac{d}{dx}\mu_{N,\downarrow}(x). \tag{11.102}$$

If there is no spin-flip at the interface, then the following boundary conditions must be satisfied:

$$\begin{aligned} J_{F,\uparrow}(x=0) &= J_{N,\uparrow}(x=0), \\ J_{F,\downarrow}(x=0) &= J_{N,\downarrow}(x=0), \\ \mu_{F,\uparrow}(x=0) &= \mu_{N,\uparrow}(x=0), \\ \mu_{F,\downarrow}(x=0) &= \mu_{N,\downarrow}(x=0). \end{aligned} \tag{11.103}$$

Furthermore, the following condition must be satisfied for all x under steady state:

$$J = J_{F,\uparrow}(x) + J_{F,\downarrow}(x) = J_{N,\uparrow}(x) + J_{N,\downarrow}(x). \tag{11.104}$$

In the set of equations (11.97 - 11.104), we have the eight unknowns (μ_1,μ_2,a_1,a_2,$c_\uparrow^F$,$c_\downarrow^F$,$c_\uparrow^N$, $c_\downarrow^N$). We can impose the condition that the total current densities in the two materials must be independent of x by setting their derivatives with respect to x equal to zero. One of the two current density expressions must be set equal to the current density flowing through the device J. This leaves us with seven equations and eight unknowns. However, a little reflection will reveal that the difference between the two unknowns μ_1 and μ_2 is an independent parameter proportional to the current density flowing through the junction. In other words, the quantities μ_1 and μ_2 can not be determined individually, but their difference is related to the bias applied between the contacts to the ferromagnet and metal.

With the use of software tools such as MATHEMATICA or MATLAB, the system of equations (11.97 - 11.104) can be solved by taking the difference $\mu_1 - \mu_2$ as a single variable. Using these results, the value of β at the interface, $\beta_I = \beta(x=0)$, is found to satisfy the following equation

$$2\beta_I - 1 = \frac{2\alpha_F - 1}{1 + 4\alpha_F(1-\alpha_F)(\sigma_N^{-1} - \lambda_{sd,N})/(\sigma_F^{-1}\lambda_{sd,F})}. \tag{11.105}$$

This last equation predicts that $\beta_I = \alpha_F = 1$ for half-metallic contact because any spin flip process can only occur on the metallic side of the junction.

Furthermore, it can be easily shown that the following inequality holds:

$$\alpha_N = 0.5 \leq \beta_I \leq \alpha_F. \tag{11.106}$$

The proof of this inequality is left to the reader as an exercise.

Introducing the weighted electrochemical potential

$$\mu_0 = \alpha(x)\mu_\uparrow(x) + (1-\alpha)\mu_\downarrow(x), \tag{11.107}$$

it can be shown that

$$\frac{d}{dx}\mu_0 = -\frac{e}{\sigma(x)}J, \tag{11.108}$$

where $\sigma = \sigma_\uparrow(x) + \sigma_\downarrow(x)$ and J is the total current density flowing through the junction. In other words $\mu_0(x)$ is the value the electrochemical potential would have had without a non-equilibrium redistribution between the up-spin and down-spin components of the current. The spatial dependence of μ_0 is shown as a tilted dashed line on either side of the junction in Fig. 11.22.

The discontinuity of μ_0 at the interface leads to a boundary resistance defined as follows

$$R_b = \frac{\Delta\mu}{eJ} = \frac{\mu_{F,0}(x=0_-) - \mu_{N,0}(x=0_+)}{eJ}, \tag{11.109}$$

where

$$\begin{aligned}\mu_{F,0}(x=0_-) &= \alpha_F\mu_{F,\uparrow}(x=0_-) + (1-\alpha_F)\mu_{F,\downarrow}(x=0_-),\\ \mu_{N,0}(x=0_+) &= \alpha_N\mu_{N,\uparrow}(x=0_+) + (1-\alpha_N)\mu_{N,\downarrow}(x=0_+).\end{aligned} \tag{11.110}$$

The results obtained using MATHEMATICA yield

$$R_b = \frac{(2\alpha_F - 1)^2(\sigma_N^{-1}\lambda_N)(\sigma_F^{-1}\lambda_F)}{(\sigma_F^{-1}\lambda_F) + 4\alpha_F(1-\alpha_F)(\sigma_N^{-1}\lambda_N)}. \tag{11.111}$$

If $\sigma_F^{-1}\lambda_F$ is much smaller than $4\alpha_F(1-\alpha_F)(\sigma_N^{-1}\lambda_N)$, then

$$R_b \sim \frac{(2\alpha_F - 1)^2\sigma_F^{-1}\lambda_F}{4\alpha_F(1-\alpha_F)}. \tag{11.112}$$

In that case, the spin current conversion mostly takes place in the ferromagnetic contact. For a half-metallic ferromagnet, $\alpha_F = 1$ and spin current conversion must occur on the metallic side, no matter the value of $\sigma_N^{-1}\lambda_N$.

Next, we use the concept of electrochemical potentials for the up-spin and down-spin channels to calculate the amount of spin polarization of the current in a semiconductor of width x_0 sandwiched between two identical ferromagnetic contacts. The efficiency of this structure as a spin valve is also analyzed.

11.7 Spin injection in a spin valve

The theory outlined in the previous section can be extended to study the efficiency of a spin valve formed of a semiconductor sandwiched between two ferromagnetic contacts separated by a distance x_0, as shown in Fig. 11.23.

The equations characterizing spin transport in all three regions assume the diffusive regime [29]:

$$\frac{d}{dx}\mu_{\uparrow,\downarrow} = -\frac{e}{\sigma_{\uparrow,\downarrow}(x)}J_{\uparrow,\downarrow}, \tag{11.113}$$

and the diffusion equation (11.94).

As in the previous section, spin flip scattering mechanisms are neglected at both interfaces with the contacts, and the electrochemical potentials and current densities in the two spin channels are assumed to be continuous throughout the structure.

Far into the bulk of both contacts, the electrochemical potentials of both spin channels must be equal. By analogy with the case of the ferromagnet/metal interface, the difference $\mu_\uparrow - \mu_\downarrow$ is also expected to reach a maximum at both interfaces and to be proportional to the total current density flowing through the structure.

According to the diffusion equation (11.94), the splitting $\mu_\uparrow - \mu_\downarrow$ decays over the spin accumulation length $\lambda_{sd} = \sqrt{D\tau_s}$ of the material (where D is the charge diffusion constant)[‖].

The spin accumulation length in the semiconductor is typically several orders of magnitude larger than the spin diffusion length in the ferromagnetic contacts [30]. Assuming an infinite spin accumulation length in the semiconductor, the diffusion equation (11.94) predicts a constant value (equal to the value at the interface) for $\mu_\uparrow - \mu_\downarrow$ inside the semiconductor. If the latter extends to infinity and the conductivities in both spin channels are equal in the semiconductor, the current densities in both channels are identical and according to Equation (11.94) no spin injection is possible, i.e., the spin-polarization of the current density $\alpha(x)$ is identically zero. However, the addition of a second ferromagnetic contact at a distance x_0 from the first one may lead to a different spatial dependence of the electrochemical potentials across the semiconductor, as illustrated qualitatively in Fig. 11.23. This is the problem that we formulate next and solve using the program MATHEMATICA. We will consider the cases where the magnetizations in the two contacts are either parallel or anti-parallel. Hereafter, any physical quantity in the left ferromagnetic contact, semiconductor channel and right ferromagnetic contact are labeled with the subscripts 1, 2 and 3, respectively.

[‖]Note that the spin accumulation length λ_{sd} and the spin diffusion length L_s will be equal only if the charge diffusion coefficient D and the spin diffusion coefficient D_s are equal. They are generally *not* equal and therefore $\lambda_{sd} \neq L_s$.

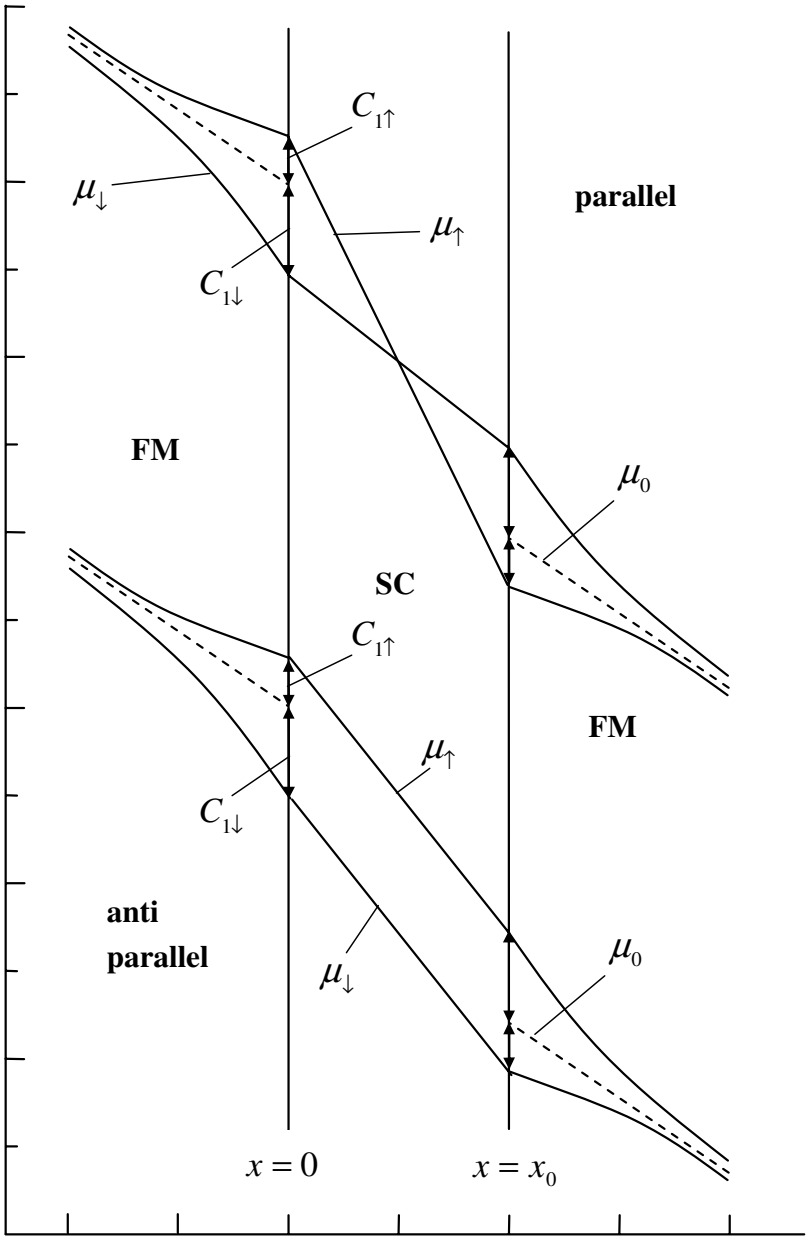


FIGURE 11.23
Illustration of the spatial dependence of the electrochemical potential in the two spin channels when the magnetizations in the identical ferromagnetic contacts are either parallel or anti-parallel. Reproduced with permission from G. Schmidt, et al., Phys. Rev. B, **62**, R4790 (2000). Copyright (2000) American Physical Society. http://link.aps.org/abstract/PRB/v62/p4790.

By analogy with Equations (11.92-11.93), the bulk spin polarization in the two ferromagnetic contacts must be such that

$$\alpha_{1,3}(\pm\infty) = \beta_{1,3} = \beta, \tag{11.114}$$

where β characterizes the fraction of the total spin current in the up-spin channel. Using the definitions (11.88) and (11.89), we then have

$$\sigma^{\uparrow}_{1,3} = \sigma_{1,3}(1 + \beta_{1,3})/2, \tag{11.115}$$

and

$$\sigma^{\downarrow}_{1,3} = \sigma_{1,3}(1 - \beta_{1,3})/2. \tag{11.116}$$

If the magnetizations of the two contacts are parallel, $\beta_1 = \beta_3$ and, if they are anti-parallel, $\beta_1 = -\beta_3$. Furthermore, if the applied bias between the two contacts is small (linear response regime), the difference in conductivity for the up-spin and the down-spin channel in the ferromagnets can easily be deduced from the Einstein relation with $D_{i\uparrow} \neq D_{i\downarrow} (i = 1, 3)$ and $\rho_{i\uparrow}(E_F) \neq \rho_{i\downarrow}(E_F) (i = 1, 3)$, where $\rho(E_F)$ is the density of states of the Fermi energy.

To separate the spin polarization effects from the normal current flow, we write the electrochemical potential in the ferromagnets for both spin directions as

$$\mu_{\uparrow,\downarrow} = \mu_0 + \mu^*_{\uparrow,\downarrow}, \tag{11.117}$$

μ_0 being the electrochemical potential without spin effects.

In the ferromagnetic contacts, we seek solutions of the diffusion equations as follows

$$\mu_{i;\uparrow,\downarrow} = \mu^0_i + \mu^*_{i\uparrow,\downarrow} = {\mu_i}^0(x) + C_{i;\uparrow,\downarrow} e^{\pm(x-x_i)/\lambda_{sd,F}}, \tag{11.118}$$

for $i = 1, 3$ with $x_1 = 0, x_3 = x_0$ and the $+(-)$ sign refers to the left and right contact, respectively.

From the boundary conditions

$$\mu_{1\uparrow}(-\infty) = \mu_{1\downarrow}(-\infty), \tag{11.119}$$

and

$$\mu_{3\uparrow}(+\infty) = \mu_{3\downarrow}(+\infty); \tag{11.120}$$

the slope of μ^0 is identical for both spin directions and also equal in regions 1 and 3 if the conductivity σ is identical in both regions.

In the semiconductor, we assume that the spin diffusion length L_s is several orders of magnitude longer than in the ferromagnet and much larger than the spacing between the two contacts and set $\tau_s = \infty$ in Equation (11.94). In this limit, the electrochemical potentials for up-spin and down-spin channels will vary linearly in the semiconductor channel

$$\mu_{2;\uparrow,\downarrow}(x) = \mu_{1;\uparrow,\downarrow}(0) + \gamma_{\uparrow,\downarrow} x, \tag{11.121}$$

where $\gamma_{\uparrow,\downarrow}$ are constants to be determined.

The last equation implies that $J_{2\uparrow}$ and $J_{2\downarrow}$ remain constant through the semiconductor. Furthermore, since we assume identical material for the two ferromagnetic contacts, the following equations must hold by symmetry

$$\mu_{1\uparrow}(0) - \mu_{1\downarrow}(0) = \pm \left[\mu_{3\downarrow}(x_0) - \mu_{3\uparrow}(x_0)\right], \tag{11.122}$$

where the $+(-)$ sign corresponds to the parallel (anti-parallel) magnetizations in the contacts, respectively.

Parallel configuration

In the left ferromagnetic contact, the spatial dependence of the electrochemical potentials in both spin channels must be of the form

$$\begin{aligned} \mu_{1\uparrow}(x) &= \mu_1^0 + \delta_{1\uparrow} x + C_{1\uparrow} e^{\frac{x}{\lambda_{sd,F}}}, \\ \mu_{1\downarrow}(x) &= \mu_1^0 + \delta_{1\downarrow} x + C_{1\downarrow} e^{\frac{x}{\lambda_{sd,F}}}. \end{aligned} \tag{11.123}$$

The corresponding spin current densities are given by [29]

$$\begin{aligned} J_{1\uparrow}(x) &= -\frac{\sigma_{F,\uparrow}}{e}\frac{\partial \mu_{1\uparrow}}{\partial x} = -\frac{1+\beta}{2}\sigma_F \left(\delta_{1\uparrow} + \frac{C_{1\uparrow}}{\lambda_F} e^{\frac{x}{\lambda_{sd,F}}}\right), \\ J_{1\downarrow}(x) &= -\frac{\sigma_{F,\downarrow}}{e}\frac{\partial \mu_{1\downarrow}}{\partial x} = -\frac{1-\beta}{2}\sigma_F \left(\delta_{1\downarrow} + \frac{C_{1\downarrow}}{\lambda_F} e^{\frac{x}{\lambda_{sd,F}}}\right). \end{aligned} \tag{11.124}$$

In the semiconductor channel, we have

$$\begin{aligned} \mu_{2\uparrow}(x) &= \mu_{1\uparrow}(0) + \gamma_\uparrow x = \mu_1^0 + C_{1\uparrow} + \gamma_\uparrow x, \\ \mu_{2\downarrow}(x) &= \mu_{1\downarrow}(0) + \gamma_\downarrow x = \mu_1^0 + C_{1\downarrow} + \gamma_\downarrow x, \end{aligned} \tag{11.125}$$

and the associated current densities are

$$\begin{aligned} J_{2\uparrow}(x) &= -\frac{\sigma_{sc,\uparrow}}{e}\frac{\partial \mu_{2\uparrow}}{\partial x} = -\frac{\sigma_{sc}}{2e}\gamma_\uparrow, \\ J_{2\downarrow}(x) &= -\frac{\sigma_{sc,\downarrow}}{e}\frac{\partial \mu_{2\downarrow}}{\partial x} = -\frac{\sigma_{sc}}{2e}\gamma_\downarrow. \end{aligned} \tag{11.126}$$

Finally, in the right contact, we have

$$\begin{aligned} \mu_{3\uparrow}(x) &= \mu_3^0 + \delta_{3\uparrow} x + C_{3\uparrow} e^{\frac{x_0 - x}{\lambda_{sd,F}}}, \\ \mu_{3\downarrow}(x) &= \mu_3^0 + \delta_{3\downarrow} x + C_{3\downarrow} e^{\frac{x_0 - x}{\lambda_{sd,F}}}, \end{aligned} \tag{11.127}$$

and

$$\begin{aligned} J_{3\uparrow}(x) &= -\frac{\sigma_{F,\uparrow}}{e}\frac{\partial \mu_{3\uparrow}}{\partial x} = -\frac{1+\beta}{2}\sigma_F \left(\delta_{3\uparrow} - \frac{C_{3\uparrow}}{\lambda_{sd,F}} e^{\frac{x_0 - x}{\lambda_{sd,F}}}\right), \\ J_{3\downarrow}(x) &= -\frac{\sigma_{F,\downarrow}}{e}\frac{\partial \mu_{3\downarrow}}{\partial x} = -\frac{1-\beta}{2}\sigma_{F,\downarrow} \left(\delta_{3\downarrow} - \frac{C_{3\downarrow}}{\lambda_{sd,F}} e^{\frac{x_0 - x}{\lambda_{sd,F}}}\right). \end{aligned} \tag{11.128}$$

Equation (11.122) leads to $C_{1\uparrow} = -C_{3\uparrow}$ and $C_{1\downarrow} = -C_{3\downarrow}$. For this case, the qualitative dependence of the electrochemical potentials is shown in the top portion of Fig. 11.23. As a result of the antisymmetric splitting of the electrochemical potentials at the interfaces, there is a different slope for $\mu_\uparrow(x)$ and $\mu_\downarrow(x)$ in the semiconductor leading to a crossing of the two electrochemical potentials half-way through the semiconductor.

Anti-parallel configuration

In this case, the down-spin (up-spin) electrons in the left contact couple to the up-spin (down-spin) electrons in the right contact, as shown schematically in the bottom part of Fig. 11.23. The electrochemical potential curves are parallel in the semiconductor channel and the total current is unpolarized. The general expressions for the electrochemical potentials and the spin current densities are still valid, but the following additional condition must be enforced: $C_{1\uparrow} = -C_{3\downarrow}$ and $C_{1\downarrow} = -C_{3\uparrow}$, with $J_\uparrow = J_\downarrow$ in the semiconductor channel. This leads to a value of α_2 exactly equal to zero.

For the parallel configuration ($\beta_1 = \beta_3$), the set of Equations (11.123-11.128) must be solved together with the following set of boundary conditions.

Boundary conditions

$$\begin{aligned}
\frac{J_{1\uparrow}(-\infty) - J_{1\downarrow}(-\infty)}{J_{1\uparrow}(-\infty) + J_{1\downarrow}(-\infty)} &= \beta, \\
\frac{J_{3\uparrow}(\infty) - J_{3\downarrow}(\infty)}{J_{3\uparrow}(\infty) + J_{3\downarrow}(\infty)} &= \beta, \\
\mu_{2\uparrow,\downarrow}(x_0) - \mu_{3\uparrow,\downarrow}(x_0) &= 0, \\
J_{1\uparrow,\downarrow}(0) - J_{2\uparrow,\downarrow}(0) &= 0 \\
J_{2\uparrow,\downarrow}(x_0) - J_{3\uparrow,\downarrow}(x_0) &= 0, \\
\frac{\partial\left(J_{1\uparrow}(x) + J_{1\downarrow}(x)\right)}{\partial x} &= 0, \\
\frac{\partial\left(J_{3\uparrow}(x) + J_{3\downarrow}(x)\right)}{\partial x} &= 0, \\
J_{1\uparrow}(0) + J_{1\downarrow}(0) &= J_{tot}.
\end{aligned} \tag{11.129}$$

Setting $\beta_1 = \beta_3$ and using MATHEMATICA or MATLAB to solve the system of equations derived above, the finite spin polarization α_2, which is independent of position in the semiconductor, can easily be calculated at $x = 0$ and is found to be equal to

$$\alpha = \beta\left(\frac{\lambda_{sd,F}}{\sigma_F}\right)\left(\frac{\sigma_F}{x_0}\right)\frac{2}{\left[2\left(\frac{\lambda_{sd,F}}{\sigma_F}\right)\left(\frac{\sigma_F}{x_0}\right) + 1 - \beta^2\right]}. \tag{11.130}$$

This expression is formally equivalent to the expression found in the resistor model equivalent circuit of the spin valve in section 11.2.3 if we make the

substitution

$$R_{fm}/R_{sc} \rightarrow \left(\frac{\lambda_{sd,F}}{\sigma_F}\right)\left(\frac{\sigma_F}{x_0}\right). \tag{11.131}$$

In the current approach, the quantities x_0/σ_F and $\lambda_{sd,F}/\sigma_F$ are proportional to the resistance of the semiconductor channel and the relevant part of the resistance of the ferromagnetic contact (which is mostly located within one spin relaxation length from each interface), respectively.

As shown in the series of exercises at the end of section 11.2.3, α_2 reaches a maximum value of β. However, this only occurs for device parameters which are virtually impossible to reach in practice, i.e., when $x_0 \rightarrow \infty$, $\sigma_{sc}/\sigma_F \rightarrow \infty$, or $\lambda_{sd,F} \rightarrow \infty$. If we use realistic values such as $\beta = 0.6$, $x_0 = 1\mu m$, $\lambda_{sd,F} = 10\ nm$ and $\sigma_F = 10^4\sigma_{sc}$, the value of α_2 is extremely small, $\sim 0.002\%$!

As an exercise, the reader is asked to reproduce figures 2 and 3 of ref. [6]. These figures show that, even for $\beta > 0.8$, $\lambda_{sd,F}$ would need to be larger than 100 nm or x_0 well below 10 nm to obtain a spin current polarization in excess of 1 %. Furthermore, when plotted as a function of β for different values of σ_F/σ_{sc}, α_2 is still below 1%, even when the $\sigma_F/\sigma_{sc} = 10$ for $\beta < 0.98$.

To assess the possibility of using a ferromagnet/semiconductor/ferromagnet junction as a spin valve, we estimate the magnetoresistance ratio $MR_1 = (R_{AP} - R_P)/R_P$ using Equation (11.133) and making the substitution (11.36). This leads to

$$MR_1 = \frac{\beta^2}{1-\beta^2}\left(\frac{\lambda_{sd,F}}{\sigma_F}\right)^2\left(\frac{\sigma_{sc}}{x_0}\right)^2\left[\frac{4}{2\left(\frac{\lambda_{sd,F}\sigma_{sc}}{\sigma_F x_0}+1\right)^2-\beta^2}\right]. \tag{11.132}$$

Exercise:

For contacts made of metallic ferromagnets, MR_1 is dominated by $\left(\frac{\lambda_{sd,F}}{\sigma_F}\right)^2\left(\frac{\sigma_{sc}}{x_0}\right)^2$. Show that in this case, $MR_1 \sim {\alpha_2}^2$.

Since α_2 is very small in practical structures, this exercise shows that MR_1 will be very difficult to detect experimentally. The prediction of small values for α_2 and MR_1 by Schmidt et al. [6] is referred to in the literature as the *conductivity mismatch* or *resistivity mismatch* problem since it is mainly due to the large difference between σ_{sc} and σ_F.

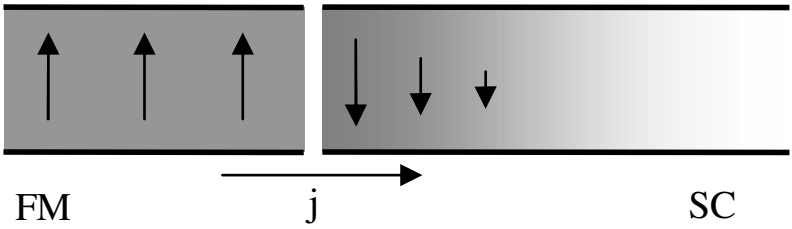


FIGURE 11.24
Illustration of the Spin Blockade phenomenon as a result of spin current extraction at a ferromagnetic contact/semiconductor interface. The flow of up-spin electrons from the semiconductor (SC) into the ferromagnet (FM) results in a higher concentration of down-spin electrons in the semiconductor channel near the interface. Reproduced with permission from Y. V. Pershin and M. Di Ventra, Phys. Rev. B, **75**, 193301 (2007). Copyright (2007) American Physical Society. http://link.aps.org/abstract/PRB/v75/p193301.

11.8 Spin extraction at the interface between a ferromagnet and a semiconductor

The study of the physics of spin extraction from a ferromagnet into a paramagnet is still in its infancy, both experimentally and theoretically [31, 32]. In this section, we present a recent analysis by Pershin and Di Ventra of electrical spin extraction from a ferromagnet into a semiconducting paramagnet in contact with it. The analysis is based on the drift-diffusion model of spin transport introduced in Chapter 10. This simple model provides a brief glimpse of the plethora of new physical phenomena to be expected in this relatively new unexplored field.

Following Pershin and DiVentra, we consider the problem of spin extraction for the contact geometry shown in Figure 11.24, in which a half-metallic ferromagnet (i.e., one with 100% spin polarization) is in contact with a paramagnetic semiconductor with a non-degenerate electron population, i.e., a population well described by Maxwell-Boltzmann statistics.

If a bias is applied across the interface such that electrons flow from the semiconductor into the ferromagnetic contact, electrons incoming from the bulk of the semiconductor are spin-unpolarized. Since the half-metallic contact accepts only up-spin electrons (the majority spins in the half-metal), a cloud of down-spin electrons must accumulate in the semiconductor close to the interface since down-spin cannot enter the ferromagnetic contact. As a result, a local spin-dipole configuration forms close to the interface. Pershin and Di Ventra refer to this phenomenon as a *Spin Blockade* which occurs at a

critical current magnitude, when the semiconductor region near the interface becomes completely depleted of electrons with spins of the same polarization as the majority spin in the half-metallic contact. In the analysis below, spin flip scattering events at the interface are neglected and the interface is modeled as a planar junction. The one-dimensional drift-diffusion equations are used to study spin extraction from the paramagnetic semiconductor.

The continuity equations are given by

$$e\frac{\partial n_{\uparrow(\downarrow)}}{\partial t} = \frac{d}{dx}J_{\uparrow(\downarrow)} + \frac{e}{2\tau_s}\left(n_{\downarrow(\uparrow)} - n_{\uparrow(\downarrow)}\right), \tag{11.133}$$

where we have introduced the spin relaxation time

$$\frac{1}{\tau_s} = \frac{1}{\tau_{\downarrow\uparrow}} + \frac{1}{\tau_{\uparrow\downarrow}}, \tag{11.134}$$

with the spin flip time $\tau_{\uparrow\downarrow}$ indicating the average time for an up-spin to flip, and $\tau_{\downarrow\uparrow}$ for the reverse process. In the paramagnetic semiconductor, we further assume that $\tau_{\uparrow\downarrow} = \tau_{\downarrow\uparrow}$.

The expressions for the current densities are

$$J_{\uparrow(\downarrow)} = \sigma_{\uparrow(\downarrow)}E_0 + eD\frac{d}{dx}n_{\uparrow(\downarrow)}, \tag{11.135}$$

where $\sigma_{\uparrow(\downarrow)}$ $(= en_{\uparrow(\downarrow)}\mu)$ are the spin conductivities and the mobility is defined such that $v_{drift} = \mu E_0$, E_0 being the assumed homogeneous electric field in the semiconductor as a result of the applied bias. Additionally, D is the diffusion coefficient in the semiconductor assumed equal for charge and spin, and also assumed equal for up-spin and down-spin electrons.

If the total electron density in the semiconductor is constant, i.e., $n_{\uparrow(\downarrow)} + n_{\downarrow(\uparrow)} = N_0$, the total current density and the homogeneous electric field are related as follows

$$J = e\mu N_0 E_0. \tag{11.136}$$

Substituting the current density expressions into the continuity equation, we obtain a set of coupled equations.

$$\frac{\partial n_{\uparrow(\downarrow)}}{\partial t} = D\frac{\partial^2 n_{\uparrow(\downarrow)}}{\partial x^2} + \mu E_0\frac{\partial n_{\uparrow(\downarrow)}}{\partial x} + \frac{n_{\downarrow(\uparrow)} - n_{\uparrow(\downarrow)}}{2\tau_s}. \tag{11.137}$$

Under steady state, we seek solutions to the previous equation of the form

$$n_{\uparrow}(x) = \frac{N_0}{2} - Ae^{-\alpha_p x}, \tag{11.138}$$

and

$$n_{\downarrow}(x) = \frac{N_0}{2} + Ae^{-\alpha_p x}. \tag{11.139}$$

Substituting the last two expressions into the steady-state version of Equation (11.133), we get a quadratic equation for α_p, whose only physically acceptable root is given by

$$\alpha_p = \frac{\mu E_0 + \sqrt{\mu^2 E_0^2 + 4\frac{D}{\tau_s}}}{2D}. \tag{11.140}$$

To find the constant A, we impose the following boundary condition far into the semiconductor

$$J_\uparrow (x \to \infty) = J_\downarrow (x \to \infty) = J/2, \tag{11.141}$$

which states that the current must be unpolarized far into the bulk. Furthermore, if we assume a half-metallic contact and full polarization at the interface, we must have

$$J_\uparrow(x = 0) = J, \tag{11.142}$$

and

$$J_\downarrow(x = 0) = 0, \tag{11.143}$$

with $x = 0$ being the location of the ferromagnet/semiconductor interface.

Since $J = e\mu \left[n_\uparrow(x) + n_\downarrow(x)\right] E_0$, it can be easily shown that Equation (11.136) is satisfied if we use the solutions (11.138) and (11.139) in the expression of the majority-spin and minority-spin current densities. Furthermore, using the expressions for the current densities to enforce the boundary conditions leads to

$$A = \frac{N_0}{\sqrt{1 + 4\frac{D}{\tau_s \mu^2 E_0^2}} - 1}. \tag{11.144}$$

This last expression increases monotonically with the total current J. The quantity A gives the deviation of $n_\uparrow(x)$ and $n_\downarrow(x)$ from their equilibrium value $N_0/2$ at the interface. It can only reach a maximum value of $N_0/2$ which occurs for a critical current density J_c given by

$$J_c = eN_0\sqrt{\frac{D}{2\tau_s}}, \tag{11.145}$$

corresponding to the maximum spin polarization (100%) or, $n_\downarrow(0) = N_0$ and $n_\uparrow(0) = 0$.

Exercises:

- For a GaAs structure, assuming that $D = D_s = 200cm^2/s$, $N_0 = 10^{15}cm^{-3}$, and $\tau_s = 10$ ns, plot α_p^{-1}, which is the inverse of the upstream spin diffusion length, as a function of E_0 for E_0 ranging 0 to 40 V/cm. What is the value of α_p^{-1} for $E_0 = 0$? Calculate the critical current density J_c in A/cm^2.

  Answer: $1/\alpha_p^{-1} = 14\ \mu m$ and $J_c = 17A/cm^2$.

  The critical current density J_c found in the exercise above is well within reach of actual experiments. This spin-blockade critical current is the maximum current allowed through the interface between a perfect (half-metallic) ferromagnet and a non-magnetic semiconductor. Your plot should show that the critical current density increases slowly as η decreases from 1 to ~ 0.3. The spin blockade phenomenon discussed above is therefore expected to be also important in junctions with ordinary ferromagnets.

  Pershin and Di Ventra recently extended the above analysis to take into account the presence of a contact resistance and calculated the current-voltage characteristics of the junction. They showed that, in the spin blockade regime, the latter saturates with increasing applied bias across the junction. A potential application of such junctions is therefore as a spin-based current stabilizer [32].

- The derivation above can be extended to the case of non-ideal ferromagnets. Defining the quantity

$$\eta = \frac{J_\uparrow(x=0) - J_\downarrow(x=0)}{J}, \tag{11.146}$$

  to characterize the level of spin polarization of the current and assuming that η does not depend on J and the ratio $J_\uparrow(x=0)/J_\downarrow(x=0)$ is a constant, prove that in this case

$$A = \frac{\eta N_0}{\sqrt{1 + 4\frac{D}{\tau_s \mu^2 E_0^2}} - 1}, \tag{11.147}$$

  and

$$J_c = eN_0\sqrt{\frac{D}{(\eta^2+\eta)\tau_s}}. \tag{11.148}$$

  Plot $J_c(\eta)/J_c(0)$ as a function of η as η varies from 1 (fully polarized spin current) to 0 (fully unpolarized spin current).

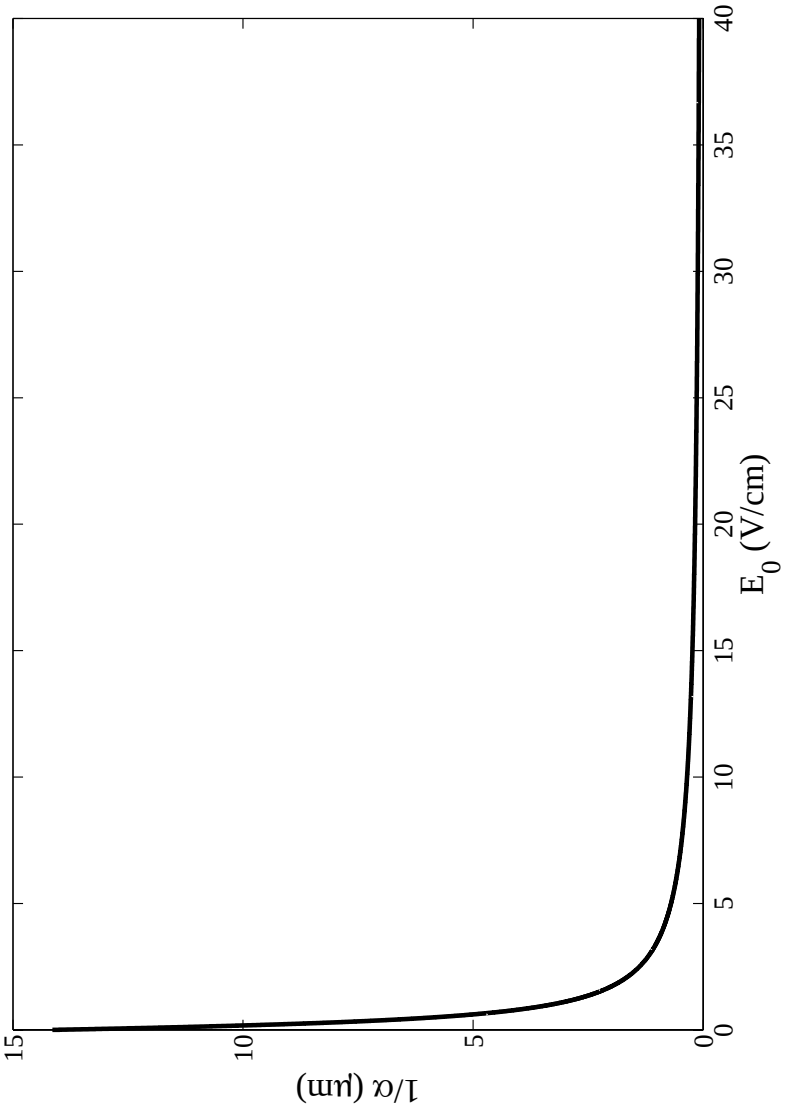


FIGURE 11.25
Plot of the up-stream spin diffusion length α_p^{-1} for a GaAs bulk sample for which D = 200 cm^2/s, $N_0 = 10^{15} cm^{-3}$ and the spin relaxation time τ_s =10 ns. The relation between the mobility and diffusion coefficients was assumed to be given by Einstein's relation and the temperature was set at 4.2 K.

11.9 Problems

- **Problem 11.1**

  Reproduce the plots shown in Figures 11.4 and 11.5 based on the theory of section 11.2.3.

- **Problem 11.2**

  Reproduce the plots shown in Figure 11.7 and Figures 11.9 through 11.12 based on the theory of section 11.2.4.

- **Problem 11.3**

  Equation (11.43) is the boundary condition expressing the discontinuity of the derivative of the wavefunction across the delta-scatterer. Show that the final expression for the transmission and reflection coefficients across the delta scatterer are the same if the boundary condition is selected to be

$$\left[\frac{d\psi}{dx}\right]_{0_+} - \left[\frac{d\psi}{dx}\right]_{0_-} = \frac{2m}{\hbar^2}\Gamma\psi(0_+). \tag{11.149}$$

  The only difference with Equation (11.43) is that the last term is selected to be $\frac{2m}{\hbar^2}\Gamma\psi(0_+)$ rather than $\frac{2m}{\hbar^2}\Gamma\psi(0_-)$. Because of the continuity of $\phi(0_+)$, the goal of this exercise is to show that this does not influence the calculation of the transmission and reflection coefficients.

- **Problem 11.4**

  Repeat the scattering problem in section 11.2.4 and derive the explicit expression of the transfer matrix across the interface at x = 0 if the scattering potential at this point is given by

$$\Gamma + \Gamma^{'} H\delta(x), \tag{11.150}$$

  where H is the Hadamard matrix.

- **Problem 11.5**

  In this problem, we consider spin transport in degenerate conductors and introduce the concept of spin-dependent electrochemical potential following Hershfield and Zhao [33]. This concept is used in section 11.6 to study the problem spin injection across a ferromagnet/non-magnetic metal interface. We only consider the case of spatial variation in one-direction (x) for simplicity. In such a system, if a small bias is applied across the junction, the equations describing the steady-state are: the

continuity equations (11.133) the current density expressions (11.135), and Poisson's equation

$$\frac{dE}{dx} = -\frac{q}{\varepsilon_0}\left(n_\uparrow + n_\downarrow\right). \tag{11.151}$$

Use those equations to prove that the up-spin and down-spin electron concentrations obey the following coupled equations (referred to as the diffusion equation)

$$\frac{d^2}{dx^2}\begin{pmatrix} n_\uparrow \\ n_\downarrow \end{pmatrix} = \begin{bmatrix} \kappa_\uparrow^2 + \frac{1}{D_\uparrow \tau_{\uparrow\downarrow}} & \kappa_\uparrow^2 - \frac{1}{D_\uparrow \tau_{\downarrow\uparrow}} \\ \kappa_\downarrow^2 - \frac{1}{D_\downarrow \tau_{\uparrow\downarrow}} & \kappa_\downarrow^2 - \frac{1}{D_\downarrow \tau_{\downarrow\uparrow}} \end{bmatrix} \begin{pmatrix} n_\uparrow \\ n_\downarrow \end{pmatrix}, \tag{11.152}$$

where

$$\kappa_\uparrow^2 = \sigma_\uparrow / \left(D_\uparrow \varepsilon_0\right), \tag{11.153}$$

and a similar expression holds for $\kappa_\downarrow^2$. The quantities $\kappa_\uparrow$ and $\kappa_\downarrow$ are referred to as the inverse up-spin and down-spin screening lengths.

Hint: Take the derivative with respect to x on both sides of equations (11.137) and use the continuity equations (11.135) and Poisson's equation.

For degenerate populations of carriers, the conductivities of the up-spin and down-spin electrons are related to their density of states and diffusion constant by Einstein's relations [34]

$$\sigma_\uparrow = e^2 g_{3D,\uparrow} D_\uparrow, \tag{11.154}$$

$$\sigma_\downarrow = e^2 g_{3D,\downarrow} D_\downarrow, \tag{11.155}$$

where $g_{3D,\uparrow}$ and $g_{3D,\downarrow}$ are the density of states in the up-spin and down-spin channels. In a non-magnetic metal, $g_{3D,\uparrow} = g_{3D,\downarrow} = \frac{1}{2} g_{3D}(E_F)$, where $g_{3D}(E_F)$ is given by Equation (11.8) in the Sommerfeld model of a metal. In a ferromagnetic contact in which the spin subbands are modeled using the Stoner-Wohlfarth model (see section 11.2.1)

$$g_{3D,\uparrow} = \frac{m_0}{2\pi^2\hbar^3}\sqrt{2m_0 E_F}, \tag{11.156}$$

and

$$g_{3D,\downarrow} = \frac{m_0}{2\pi^2\hbar^3}\sqrt{2m_0\left(E_F - \Delta\right)}, \tag{11.157}$$

where E_F and Δ are the Fermi and exchange splitting energy of the ferromagnet, respectively.

- **Problem 11.6**

  Following Hershfield and Zhao [33], we introduce the electrochemical potentials $\mu_\uparrow(x)$ and $\mu_\downarrow(x)$ for the up-spin and down-spin electrons such that their corresponding spin concentrations are given by

$$n_\uparrow(x) = \rho_{3D,\uparrow}(E_F)\left[\mu_\uparrow(x) + eV(x)\right], \tag{11.158}$$

  and

$$n_\downarrow(x) = \rho_{3D,\downarrow}(E_F)\left[\mu_\downarrow(x) + eV(x)\right], \tag{11.159}$$

  where $V(x)$ is the electrostatic potential.

  Using the last two equations, show that the total current densities in Equations (11.135) can be written in the more compact form

$$J_\uparrow(x) = \frac{\sigma_\uparrow}{e}\frac{d\mu_\uparrow}{dx}, \tag{11.160}$$

  and

$$J_\downarrow(x) = \frac{\sigma_\downarrow}{e}\frac{d\mu_\downarrow}{dx}, \tag{11.161}$$

  Solution: Starting with Equation (11.159), we get

$$\frac{\mu_\uparrow(x)}{-e} = V(x) + \frac{n_\uparrow(x)}{eg_{3D,\uparrow}(E_F)} \tag{11.162}$$

  and a similar expression for $\mu_\downarrow(x)/-e$.

  Taking the derivative with respect to x on both sides of the last equation, we get

$$\frac{d}{dx}\left(\frac{\mu_\uparrow(x)}{-e}\right) = \frac{dV(x)}{dx} + \frac{1}{eg_{3D,\uparrow}(E_F)}\frac{dn_\uparrow(x)}{dx}. \tag{11.163}$$

  Multiplying the last equation on both sides $-\sigma_\uparrow$ and taking into account that the electric field $E(x) = -dV/dx$ and Einstein's relations (11.154,11.155), we get

$$J_\uparrow = -\sigma_\uparrow\frac{d}{dx}\left(\frac{\mu_\uparrow(x)}{-e}\right) = \sigma_\uparrow E + qD_\uparrow\frac{dn_\uparrow}{dx}, \tag{11.164}$$

and a similar expression for $J_\downarrow$.

The spin-dependent electrochemical potentials are useful quantities since they allow a direct calculation of the up-spin and down-spin electron concentrations and simple spatial derivatives give the associated spin current densities.

- **Problem 11.7**

  Starting with the results of the previous problem, show that the electrochemical potentials obey the following set of coupled equations

$$\frac{d^2}{dx^2}\begin{pmatrix}\mu_\uparrow\\ \mu_\downarrow\end{pmatrix} = \begin{pmatrix}+\frac{1}{D_\uparrow\tau_{\uparrow\downarrow}} & -\frac{1}{D_\uparrow\tau_{\uparrow\downarrow}}\\ -\frac{1}{D_\downarrow\tau_{\downarrow\uparrow}} & +\frac{1}{D_\downarrow\tau_{\downarrow\uparrow}}\end{pmatrix}\begin{pmatrix}\mu_\uparrow\\ \mu_\downarrow\end{pmatrix}. \tag{11.165}$$

  If we assume $\tau_{\downarrow\uparrow} = \tau_{\uparrow\downarrow} = \tau_s$, the spin flip scattering rate, show that the difference of the spin-dependent electrochemical potential $\mu_\uparrow - \mu_\downarrow$ obey the diffusion equation

$$\frac{d^2}{dx^2}\left(\mu_\uparrow - \mu_\downarrow\right) = \frac{\mu_\uparrow - \mu_\downarrow}{L_{sf}^2}, \tag{11.166}$$

  where the spin decay length is given by $L_{sf} = \sqrt{D_{eff}\tau_s}$ and D_{eff} is the effective diffusion coefficient and is equal to

$$D_{eff} = \frac{D_\uparrow D_\downarrow}{D_\uparrow + D_\downarrow}. \tag{11.167}$$

11.10 References

[1] N.F. Mott, "The electrical conductivity of transitions metals", Proc. R. Soc. London, Ser. A, **156**, 699 (1936); "The resistance and thermoelectric properties of the transition metals", Proc. R. Soc. London, Ser. A, **156**, 368 (1936).

[2] N. W. Ashcroft and N. D. Mermin, *Solid State Physics*, (Saunders College, Philadelphia, 1976).

[3] M. Jullière, "Tunneling between ferromagnetic films", Phys. Lett. A, **54**, 225 (1975).

[4] M. N. Baibich, J. M. Broto, A. Fert, F. Nguyen Van Dau, and F. Petroff, P. Eitenne, G. Creuzet, A. Friederich and J. Chazelas, "Giant magnetoresistance of (001)Fe/(001)Cr magnetic superlattices", Phys. Rev. Lett., **61**, 2472 (1988).

[5] G. Binasch, P. Grünberg, F. Saurenbach and W. Zinn, "Enhanced magnetoresistance in layered magnetic structures with antiferromagnetic interlayer exchange", Phys. Rev. B, **39**, 4828 (1989).

[6] G. Schmidt, D. Ferrand, L. W. Molenkamp, A. T. Filip and B. J. van Wees, "Fundamental obstacle for electrical spin injection from ferromagnetic metal into a diffusive semiconductor", Phys. Rev. B, **62**, R4790 (2000).

[7] E. I. Rashba, "Theory of electrical spin injection: Tunnel contacts as a solution of the conductivity mismatch problem", Phys. Rev. B, **62**, R16267 (2000).

[8] R. Feiderling, M. Klein, G. Reuscher, W. Ossau, G. Scmidt, A. Waag and L. Molemkamp, "Injection and detection of a spin-polarized current in a light-emitting diode", Nature (London), **402**, 787 (1999).

[9] G. Salis, R. Wang, X. Jiang, R.M. Shelby, S.S.P. Parkin, S.R. Bank and J.S. Harris, "Temperature independence of the spin injection efficiency of a MgO based tunnel spin injector", Appl. Phys. Lett., **87**, 262503 (2005).

[10] T. Santos, J. Lee, P. Migdal, I. Lekshmi, B. Satpati and J. Moodera, "Room temperature tunnel magnetoresistance and spin polarized tunneling through an organic semiconductor barrier", Phys. Rev. Lett., **98**, 016601 (2007).

[11] W.E. Pickett and J.S. Moodera, "Half metallic magnets", Physics Today, **54**, 39 (2001).

[12] P.A. Dowben and R. Skomski, "Are half metallic ferromagnets half metals?", J. Appl. Phys., **95**, 7453 (2004).

[13] M. S. Lundstrom, *Fundamentals of Carrier Transport*, Purdue Modular Series in Solid State Devices, Eds. G. W. Neudeck and R. F. Pierret, (Addison-Wesley, Reading, MA, 1990).

[14] A. D. Stone and A. Szafer, "What is measured when you measure a resistance? The Landauer formula revisited", IBM J. Res. Develop., **32**, 384 (1988).

[15] J. Bardeen, "Tunneling from a many-particle point of view", Phys. Rev. Lett., **6**, 57 (1961).

[16] S. J. Xie, K. H. Ahn, D. L. Smith, A. R. Bishop and A. Saxena, "Ground state properties of ferromagnetic metal/cojugated polymer interfaces", Phys. Rev. B, **67**, 125202 (2003).

[17] K. Hathaway and E. Dan Dahlberg, "Resource Letter STMN-1: Spin Transport in magnetic nanostructures", Am. J. Phys., **75**, 871 (2007).

[18] J. Mathon, "Phenomenological theory of giant magnetoresistance", Chapter 4 in *Spin Electronics*, edited by M. Ziese and M.J. Thornton, (Springer Verlag, Berlin, 2001).

[19] A. Barthélémy, A. Ferta, J.-P. Contour, M. Bowen, V. Cros, J.M. De Teresa, A. Hamzic, J.C. Faini, J.M. George, J. Grollier, F. Montaigne, F. Pailleux, F. Petroff and C. Vouille, "Magnetoresistance and spin electronics", Journal of Magn. and Magn. Mater., Vol.242-245, 68 (2002).

[20] M.A.M. Gijs and G.E.W. Bauer, "Perpendicular giant magnetoresistance of magnetic multilayers", Adv. Phys., **46**, 285 (1997).

[21] J.M. Daughton, "Magnetic tunneling applied to memory", J. Appl. Phys., **81**, 3758 (1997).

[22] J.M. Daughton, A.V. Pohm, R.T. Fayfield and C.H. Smith, "Applications of spin dependent transport materials", J. Phys. D, **32**, R169-R177 (1999).

[23] J.M. Daughton, J. Brown, E. Chen, R. Beech, A.V. Pohm and W. Kude, "Magnetic field sensors using GMR multilayer", IEEE Trans. Mag., **30**, 4608 (1994).

[24] G.A. Prinz, "Magnetoelectronics", Science, **282**, 1660 (1998).

[25] J. Hayakawa, S. Ikeda, Y. M. Lee, F. Matsukura and H. Ohno, "Effect of high annealing temperature on giant tunneling magnetoresistance ratio of CoFeB/MgO/CoFeB magnetic tunnel junctions", Appl. Phys. Lett., **89**, 232510 (2006).

[26] World magnetic sensor components. Frost & Sullivan (2005).

[27] A. Fert, J.-L. Duvail and T. Valet, "Spin relaxation effects in the perpendicular magnetoresistance of magnetic layers", Phys. Rev. B, **52**, 6513 (1995).

[28] P.C. van Son, N. van Kampen and P. Wyder, "Boundary resistance of the ferromagnetic-nonferromagnetic metal interface", Phys. Rev. Lett., **58**, 2271 (1987).

[29] In Equations (99) - (102), we introduce a negative sign so that the current densities associated with the flow of electrons from left to right in Fig. 11.22 are positive. The same convention is used in Equations (113), (124), (126) and (128).

[30] Spin diffusion length approaching 100 μm have been reported by J.M. Kikkawa and D.D. Awschalom, "Lateral drag of spin coherence in gallium arsenide", Nature (London), **397** (1999).

[31] Y. V. Pershin and M. Di Ventra, "Spin blockade at semiconductor/ferromagnet junctions", Phys. Rev. B, **75**, 193301 (2007).

[32] Y. V. Pershin and M. Di Ventra, "Current-voltage characteristics of semiconductor/ferromagnet junctions in the spin blockade regime", arXiv:0707.4475v1 (2007).

[33] S. Hershfield and H.L. Zhao, "Charge and spin transport through a metallic ferromagnetic-paramagnetic-ferromagnetic junction", Phys. Rev. B, **56**, 3296 (1997).

[34] S. Datta, *Electronic Transport in Mesoscopic Systems*, (Cambridge University Press, New York, 1995), p. 41.

12

Hybrid Spintronics: Active Devices Based on Spin and Charge

In Chapter 11, we discussed passive spintronic devices. In this Chapter, we introduce and discuss *active* spin-based devices such as spin transistors, that are capable of amplifying a signal. They can be used in signal processing and logic circuits. But first, we must make an important distinction between what we call "hybrid spintronics" and "monolithic spintronics".

We mentioned in the preface of this textbook that "spintronics" is the science and technology of utilizing the spin degree of freedom of a charge carrier (electron or hole) to store, process and communicate information. Both classical and quantum mechanical information can be encoded and manipulated by spin; however, spin's role in all this can be either central or peripheral. That brings us to a discussion of *hybrid* and *monolithic* spintronics. The distinction between the two is that in the former case, spin does not directly store or process information. That role is still delegated to charge, but spin influences how charge stores or manipulates information. Therefore, spin has a secondary role and may, at best, augment the role of charge. Spin-based transistors, which are discussed in this Chapter, are hybrid devices in this sense. Monolithic spintronics, on the other hand, deals with devices in which spin single handedly stores and processes information, without involving charge at all. The quintessential example of monolithic spintronics is Single Spin Logic (SSL) which is discussed in the next Chapter. Fig. 12.1 shows the hierarchy of spintronics.

12.1 Spin-based transistors

The basic definition of a "transistor" is a three-terminal device where the current flowing between two of the terminals can be modulated by a current or voltage applied to the third terminal. Device engineers would cringe at this simplistic definition of a transistor, since they will usually recognize that the transistor is much more than just a three-terminal device with current modulation. The current modulation is a necessary, but *not* a sufficient, condition for a device to qualify as a transistor. A true transistor must have

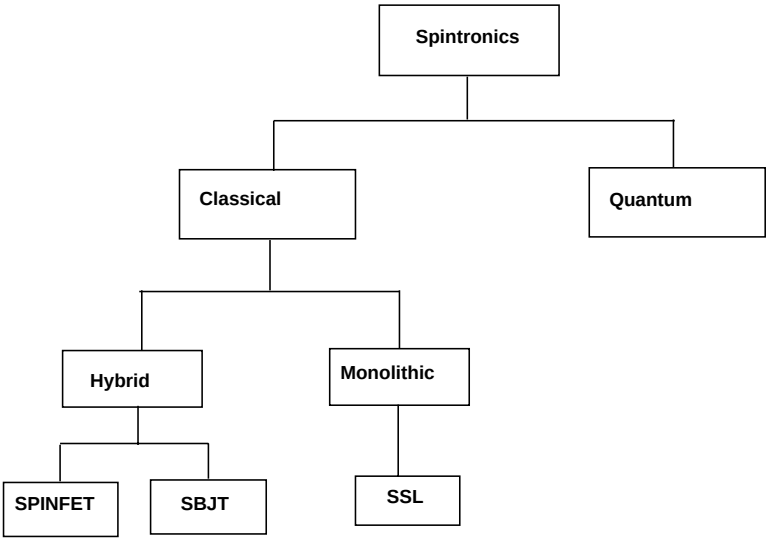


FIGURE 12.1
The hierarchy of spintronics.

other attributes as well, such as "gain" and "isolation between the input and output terminals". The "gain" (or amplification) requirement means that a small change in the current or potential at the third terminal (the input signal) should produce a much larger change in the current flowing between the other two terminals or the potential difference appearing between those terminals (the output signal). "Isolation" is an even more important property. It means that any change in current or potential at the input terminals should bring about a change in the current or potential at the output terminal, but this should be a *one-way* interaction. In other words, any change caused at the output terminal by external means should not induce a change in the input quantity. Stated otherwise, signal should be communicated "unidirectionally" from the input to the output and not the other way around. Both "gain" and "unidirectionality" (or "isolation") are required for digital logic circuits [1]. As we embark on a discussion of spin transistors, the reader will do well to keep these ideas in mind for reasons that will become clear later.

In the following sections, we will discuss two genres of spin-based transistors: Spin Field Effect Transistors and Spin Bipolar Transistors.

12.2 Spin field effect transistors (SPINFET)

In many ways, the first proposal for a Spin Field Effect Transistor (SPINFET) [2] is a watershed event in the field of spintronics. It was the first time anyone proposed using the spin degree of freedom of a charge carrier to realize an active device which can process information in a tractable way. The SPINFET device proposed in ref. [2] looks exactly like a conventional metal-insulator-semiconductor field effect transistor (MISFET) except that the source and drain contacts are ferromagnetic. Figure 12.2 shows a schematic of this device. To keep the ensuing discussion focused on the essential elements, we will assume that the channel is strictly one dimensional (a quantum wire) whose width is smaller than the Fermi wavelength of carriers in the channel, so that only the lowest subband is occupied by electrons.

Both source and drain contacts are magnetized in such a way that their magnetic moments are parallel to each other and point along the direction of current flow (+x-axis). When a potential difference is imposed between these contacts, the ferromagnetic source injects carriers into the channel with their spins polarized in the +x-direction. These are the majority spins in the ferromagnetic source contact. We will now make four idealized assumptions:

1. The ferromagnetic source injects only the majority spins and no minority spins at all. Moreover, there should be no loss of spin polarization at the interface due to spin flip scattering. In other words, the spin injection

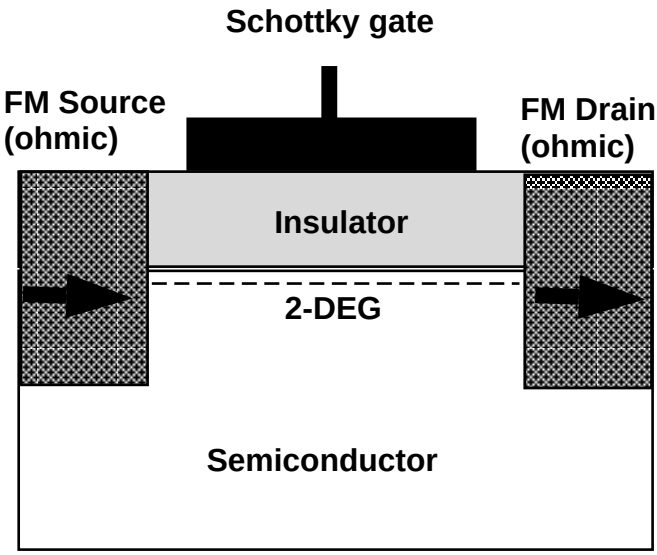


FIGURE 12.2
Schematic representation of a Spin Field Effect Transistor. The arrows indicate the direction of magnetization of the source and drain contacts.

efficiency at the source end, defined as

$$\zeta = \frac{I_{maj} - I_{min}}{I_{maj} + I_{min}}, \tag{12.1}$$

is 100%, where I_{maj} is the current in the channel due to majority spins (in the source) and I_{min} is the current due to minority spins.

2. We will also assume that the drain is an idealized spin filtering ferromagnet which only transmits the majority spins and completely blocks the minority spins.

3. When the gate voltage is zero, there are no stray symmetry breaking electric fields in the y-z plane that induce any Rashba interaction. Note that the source-to-drain electric field, that drives the current in the channel, is in the x-direction which is the same direction as the carrier velocity; hence, it does not cause a Rashba spin-orbit interaction. Only when the gate voltage is non-zero can there be a Rashba interaction in the channel due to the electric field caused by the gate voltage. There is also no Dresselhaus spin-orbit interaction, and no stray magnetic fields in the channel (we ignore the magnetic field caused by the ferromagnetic contacts).

4. There is no spin relaxation in the channel. There can be momentum and energy relaxing collisions in the channel – transport need not be ballistic – but these scatterings must not relax spin. This is a realistic assumption since we know (see Chapter 8) that in a strictly one dimensional channel with a single transverse subband occupied, there is no D'yakonov-Perel' spin relaxation in space. There is also no Elliott-Yafet spin relaxation if there is no magnetic field in the channel. This is because if the only spin-orbit interaction is the Rashba interaction, then each spin-split band has a fixed (wavevector-independent) spin quantization axis (see Section 7.2.1.). In that case, intraband scattering does not change spin polarization and interband scattering (due to a non-magnetic scatterer) is forbidden because the spin eigenstates in two different bands are orthogonal. Therefore, there is no Elliott-Yafet relaxation.

Right after injection, all spins are polarized along the direction of current flow, i.e., +x-direction. When the gate voltage is turned on, it causes a transverse electric field in the y-direction, which causes Rashba spin-orbit interaction in the channel. This interaction will act as an effective magnetic field in the z direction which will be given by (see Equation (8.18))

$$B_{Rashba} = \left[\frac{2m^* a_{46}}{g\mu_B \hbar}\mathcal{E}_y v_x\right]\hat{z}\ , \tag{12.2}$$

where $\mathcal{E}_y$ is the y-directed gate electric field and v_x is the x-directed velocity of carriers in the channel. Note that we have used the spin-independent field which is proportional to velocity and not wavevector. This will allow us to work with a spin-independent precession frequency Ω (see Problem 8.5).

The pseudo-magnetic field B_{Rashba} is directed along the z-axis. It will cause the electrons entering the channel with x-polarized spins to execute Larmor precession in the x-y plane as they travel towards the drain. The angular frequency of this precession is the Larmor frequency

$$\Omega = \frac{d\phi}{dt} = \frac{g\mu_B B_{Rashba}}{\hbar} = \frac{2a_{46}m^*}{\hbar^2}\mathcal{E}_y v_x, \tag{12.3}$$

which is **spin-independent** since we have used the spin-independent B_{Rashba}.

Therefore, the *spatial rate* of spin precession will be given by

$$\frac{d\phi}{dx} = \frac{d\phi}{dt}\frac{dt}{dx} = \frac{d\phi}{dt}\frac{1}{v_x} = \frac{2a_{46}m^*}{\hbar^2}\mathcal{E}_y. \tag{12.4}$$

Needless to say, this is also spin-independent.

It is important to note from the above equation that the spatial rate of spin precession depends on the gate voltage (or $\mathcal{E}_y$) but is *independent* of the carrier velocity along the channel. This means that the spin of every electron, regardless of its velocity, precesses by exactly the *same* angle as it travels from source to drain. This angle is given by

$$\Phi = \left[\frac{2a_{46}m^*}{\hbar^2}\mathcal{E}_y\right] L \ , \tag{12.5}$$

where L is the length of the channel (or source-to-drain separation).

An electron may suffer momentum randomizing collision and arrive at the drain with arbitrary velocity, but it does *not* matter. When *any* electron arrives at the drain, with whatever velocity and whatever scattering history, its spin has precessed by exactly the same angle Φ as any other electron in traversing the channel*. The angle Φ depends only on $\mathcal{E}_y$, or the gate voltage, which is the same for every electron. If the gate voltage is of such magnitude that $\Phi = (2n+1)\pi$, then *every* electron arriving at the drain will have its spin polarized *anti-parallel* to the drain's magnetization, since they all entered the channel with spins polarized parallel to the drain's magnetization. These electrons are blocked by the drain since the drain transmits only those

*The reader should appreciate the fact that it is absolutely necessary that B_{Rashba} be proportional to the carrier velocity and *not* the carrier wavevector (which is not proportional to the velocity), for this discussion to be valid. This is a matter of consistency. If we assume a spin-independent precession frequency Ω, as we have done here, we must also invoke a spin-independent spin-orbit field B_{Rashba} since Ω is proportional to B_{Rashba}. We have seen earlier that if B_{Rashba} is proportional to v then it is spin-independent, but if it is proportional to k then it is not. Hence, the B_{Rashba} used in this discussion must be proportional to v.

electrons whose spins are majority spins in the drain ferromagnet. Therefore the source to drain current falls to zero. Without a gate voltage, the spins do not precess ($\Phi = 0$ when $\mathcal{E}_y = 0$) so that every electron arriving at the drain has its spin polarized *parallel* to the drain's magnetization. These electrons are all transmitted by the drain so that the source-to-drain current is non-zero. Thus, the gate voltage causes current modulation via spin precession and realizes basic transistor action. This is the basis of the celebrated *Spin Field Effect Transistor*.

Datta and Das [2] considered a ballistic device. For such a device, realized with a quantum wire where only the lowest subband is occupied, the source to drain conductance G_{SD} (assuming linear response) can be calculated from the Landauer formula. The result is

$$G_{SD} = \frac{e^2}{h} cos^2\left(\frac{\Phi}{2}\right), \tag{12.6}$$

where we have ignored any reflection at the contact interfaces. Since Φ depends on the gate voltage, the conductance G_{SD} can be modulated with the gate voltage V_G. In Fig. 12.3 we plot the transfer characteristic (source-to-drain current I_{SD} versus gate voltage V_G) of an ideal SPINFET. Note that this device is "normally on", meaning that the current I_{SD} is maximum when the gate voltage V_G is zero and a gate voltage V_{OFF} is required to turn the device off. The transfer characteristic is *oscillatory*. Therefore, the transconductance $\partial I_{SD}/\partial V_G$) can be either positive or negative depending on the dc bias applied to the gate terminal. If two devices – one with positive transconductance and one with negative transconductance – are connected in series, then we will realize a device that acts like a complementary metal oxide semiconductor (CMOS) inverter. Oscillatory transfer characteristics have other unusual applications. Suppose we apply a sinusoidal voltage of frequency f to the gate and the amplitude of this voltage is n times V_{OFF}. Then the source to drain current I_{SD} will oscillate with a frequency nf. Thus, we have realized a single stage frequency multiplier.

It should be obvious that in this device, although current modulation is achieved through spin precession, spin itself plays no direct role in information handling. Information is still encoded in charge which carries the current from the source to the drain. The transistor is still switched between the on and off states by changing the current or the amount of charge transmitting through the device. The role of spin is only to provide an alternate means of changing the current. Thus, this device is a quintessential hybrid spintronic device.

One final remark about the Datta-Das SPINFET is in order. There is a misconception among some that this device is an "interference device", implying that some kind of phase coherence is required. In reality, no phase coherence is required and the device is completely *classical*. Its operation can be explained (at least in the one-dimensional case) by invoking classical Larmor precession of spin about an effective velocity-dependent magnetic field B_{Rashba} caused by the Rashba interaction. Datta and Das analyzed their

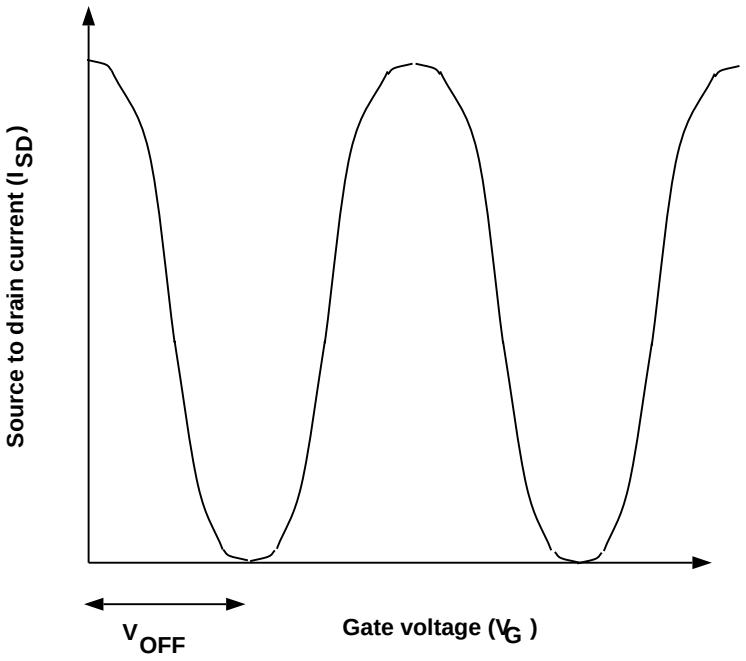


FIGURE 12.3
The transfer characteristic (source-to-drain current versus gate voltage) of an ideal SPINFET. The gate voltage required to turn the transistor from 'on' to 'off' is V_{OFF}. Note that the device is "nomrally-on" since the source-to-drain current is maximum when the gate voltage is zero.

device in a different way without invoking the Larmor precession picture. We used the Larmor precession picture purposely to highlight the classical nature of the SPINFET. This device can tolerate momentum randomizing and phase breaking collisions of electrons in the channel as long as they do not flip spin. That makes it much more robust than any quantum interference device.

Significant effort has been made to demonstrate this device experimentally, but at the time of writing this textbook, no effort has borne fruit. The primary obstacles to realizing this device is that spin injection and detection efficiencies at ferromagnet/semiconductor interfaces are much smaller than 100% contrary to what assumptions 1 and 2 assumed. If this technological barrier can be overcome, the Datta-Das SPINFET might indeed be demonstrated some day. Whether it will lead to a better device, superior to traditional transistors, is another issue that we discuss later in this chapter.

12.2.1 Sub-threshold slope

One way the Datta-Das device is different from a conventional field effect transistor (FET) is that, in principle, the average carrier concentration in the channel need not change at all when the device is switched from "on" (conducting) to "off" (non-conducting), or vice versa[†]. In conventional n-channel FETs, the current flowing between the source and drain is changed by altering the electron concentration in the channel with the gate voltage. If the FET is a so-called "enhancement mode" (or "normally-off") device, then without any applied gate voltage, there should be no (or very few) electrons in the channel. When a positive voltage is applied to the gate, it will draw electrons in the channel from the source contact by Coulomb attraction, thereby increasing the channel conductance and the channel current for a fixed source-to-drain voltage. By contrast, the channel concentration in a SPINFET need not change; it is only the transmittivity of the carriers that changes since the transmittivity depends on the spin polarization. This mode of switching, where the channel carrier concentration does not change, may have some advantages. Consider the conventional FET in the "off" state. The conduction band energy profile along the length of the channel is shown in Figure 12.4. The channel is depleted of electrons, as it should be in the OFF-state, and the Fermi level (determined by the Fermi level in the metallic contacts) is well below the conduction band edge, consistent with the fact that the channel is empty of carriers. When a positive gate voltage is applied to turn the transistor "on", it lowers the conduction band edge to a position below the Fermi level. The barrier at the source-channel interface is also made *thinner*, so that electrons get into the channel by tunneling through this barrier from the

[†]In reality, the gate voltage will always cause some change in the channel carrier concentration since the charges on the gate will attract or repel charge carriers in the channel. However, this change could be very small if a very small gate voltage is required to turn the device on or off, or if the gate capacitance is very small.

source. Since the conduction band edge is below the Fermi level, the channel becomes populated.

In the "off' state, the barrier is too thick to tunnel through, but some carriers can still get into the channel by thermionic emission over the barrier. These electrons cause a leakage current in the "off" state that leads to unwanted standby power dissipation in the device. Assuming Boltzmann statistics for the electrons, the population in the channel will fall off as $exp[-\phi_b/kT]$ where ϕ_b is the effective barrier height. We will assume that ϕ_b is proportional to eV_G, where V_G is the gate voltage. Therefore the leakage current will fall off with a slope of $(kT/e)ln(10)$ per decade of gate voltage (assuming an enhancement mode device). This slope is called the sub-threshold slope and has a minimum value of $(kT/e)ln(10)$/decade = 60 mV/decade at room temperature. It is the *minimum* value because we have completely neglected tunneling in the off state.

It should be obvious that the 60 mV/decade (at room temperature) limitation comes about because we switch the transistor on and off by populating and de-populating the channel. Instead, if we can switch the device by changing the carrier transmittivity with a gate voltage, without changing the carrier concentration, then the 60 mV/decade limit need not apply. This could have been one advantage of the SPINFET[‡]. However, we will show in the next subsection that, realistically speaking, a SPINFET will probably have a much larger leakage current than a conventional FET for entirely different reasons.

One last issue that we need to address before we embark on a realistic assessment of the device potential of SPINFETs is the issue of "normally on" versus "normally off". It is obvious that the SPINFET, as described, is "normally on", meaning that it is in the conducting state ("on") when the gate voltage is zero. This is not desirable. If we want to build the more desirable "normally off" device, we will simply have to magnetize the source and drain ferromagnets in the anti-parallel configuration so that the majority spins injected by the source are minority spins in the drain, and vice versa. We need not actually magnetize the two contacts in opposite directions so that their magnetic moments are pointing against each other (the 'north pole' of

[‡]In the 1980s, a device known as *Velocity Modulation Transistor* became popular (H. Sakaki, Jpn. J. Appl. Phys., Pt. 2, Vol. 21, L381 (1982) and C. Hamaguchi, Jpn. J. Appl. Phys., Pt. 2, Vol. 23, L132 (1984)). The device structure was that of a MOSFET and the channel was a quantum well with a fixed carrier concentration. A positive gate voltage pulled the electrons towards an interface of the quantum well, thereby increasing interface roughness scattering in the channel and decreasing carrier mobility. This decreased the conductance of the channel, realizing transistor action. Just like the SPINFET, this device is not subject to the $(kT/e)ln(10)$/decade sub-threshold slope limitation. In fact, any device where the carrier concentration is not modulated for transistor action will not be subjected to this limitation. It turned out later that the on-off conductance ratio of the Velocity Modulation Transistor was too small for mainstream applications since mobility could not be changed by several orders of magnitude using this technique. As a result, the Velocity Modulation Transistor idea never caught on. We will show later that SPINFETs will also tend to have a very small on-to-off conductance ratio.

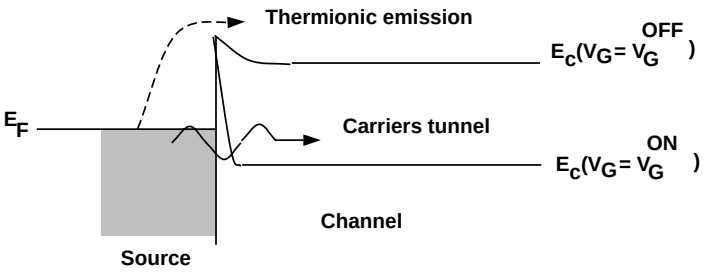


FIGURE 12.4
Conduction band energy profile in a MISFET along the length of the channel (axis joining source and drain). When the gate voltage is such that the transistor is OFF, the barrier at the source/channel interface is wide. That prevents carriers from tunneling into the channel. When the gate voltage turns the transistor ON, the conduction band edge is lowered and barrier made thin. This allows carriers to tunnel into and populate the channel. Thermionic emission can inject carriers into the channel even when the device is OFF.

one magnet facing the 'north pole' of the other). Such a configuration will be unstable and both source and drain ferromagnets will demagnetize quickly. A better alternative is to choose two different materials, such as iron and cobalt, which have opposite signs of spin polarization of carriers at the Fermi level [3] so that majority spins in one are already minority spins in the other when both are magnetized in the same direction. A SPINFET with iron and cobalt contacts will be normally off. "Normally off" devices are always preferred to avoid standby power dissipation in circuits.

12.2.2 The effect of non-idealities

Spin injection efficiency: Let us now re-examine some of the assumptions that we made when we presented the theory of the ideal SPINFET in Section 12.2. The first assumption was that the *spin injection efficiency* at the source-channel interface and the *spin detection efficiency* at the drain-channel interface are both 100%. At the time of writing this Chapter, the highest electrical spin injection efficiency demonstrated at a ferromagnet/paramagnet interface was only 70% at or near room temperature [4]. Perfect 100% spin injection efficiency would require that the injecting ferromagnet is an ideal half metal with 100% spin polarization [5]. Likewise, perfect 100% spin detection efficiency would require that the detecting ferromagnet is an ideal ferromagnet

with 100% spin polarization. It has been argued that ideal half metals cannot exist at any temperature above 0 K because of magnons and phonons, and even at 0 K, surfaces and inhomogeneities will reduce the spin polarization to less than 100% [6]. Thus, the spin injection and detection efficiencies are probably going to be less than 100%.

We will show later that the ratio of "on-conductance" to "off-conductance" of the SPINFET has a maximum value of $G_{on}/G_{off} = (1+\zeta_S\zeta_D)/(1-\zeta_S\zeta_D)$, where ζ_S is the spin injection efficiency at the source and ζ_D is the spin detection efficiency at the drain. If $\zeta_S = \zeta_D = 70\%$, then this ratio has a value of 2.92, which is clearly inadequate for any mainstream application. This small ratio means that the leakage current in the off state is huge; it is at least one-third of the on-current, if spin injection/detection efficiencies are 70%. In order to achieve a conductance on-off ratio of 10^5, required of modern transistors, $\zeta_S = \zeta_D = 99.9995\%$, which is a very tall order. The *primary impediment* to realization of usable SPINFETs at this time is inadequate spin injection efficiency. We will re-visit this later.

Channel magnetic field: Another questionable assumption is the absence of stray magnetic fields in the channel. Invariably, there will be a magnetic field along the channel because of the magnetized ferromagnetic contacts. This will cause problems. The first problem is that the spin precession rate in space will be no longer given by Equation (12.4) and will be no longer independent of the carrier velocity (or energy). This is because the spins will precess about the net magnetic field, and this net field is

$$\vec{B}_{total} = \vec{B}_{Rashba} + \vec{B}_{channel} \tag{12.7}$$

where the first term on the right hand side is linearly proportional to the carrier velocity v_x, but the second term is independent of velocity. Therefore, the spatial rate of spin precession, which is $d\phi/dx = g\mu_B B_{total}/(\hbar v_x)$, is no longer independent of carrier velocity v_x. Consequently, the angle by which an electron's spin has precessed when it arrives at the drain contact (Φ) will depend on the electron's velocity and scattering history. At a finite temperature, and in the presence of momentum randomizing scattering in the channel, the spread in the electron velocity will make different electrons suffer different amounts of spin precession Φ, so that when they arrive at the drain end, not all of them will have the same spin polarization. Those who have their spin polarizations anti-parallel to the drain's magnetization will be blocked, but the others will transmit to varying degrees and cause a leakage current in the off-state. This is another source of leakage current, although it is less important than that due to imperfect spin injection and detection at the source and drain contacts. A more serious problem is that the channel magnetic field makes the eigenspinors wavevector dependent (recall Chapter 7). Without the magnetic field, the Rashba interaction lifts the spin degeneracy at any non-zero wavevector, but each spin-split band still has a fixed spin quantization axis that is parallel or anti-parallel to B_{Rashba}.

This means that the spin polarization in each band is always the same and independent of wavevector. Furthermore, the spin polarizations in the two bands are anti-parallel to each other. Thus, there can be no scattering between the two bands because their eigenspinors are orthogonal. Electrons can scatter elastically or inelastically only within the same spin band, but not between two different spin bands. Intraband scattering does not alter the spin polarization since every state in the same band has exactly the same spin polarization. Therefore, scattering caused by any non-magnetic entity (phonons, impurities, etc.) does *not* relax spin, if there is no magnetic field in the channel. In other words, Elliott-Yafet type spin relaxation is not possible in a quantum wire channel if there is no channel magnetic field. We also know that there can be no D'yakonov-Perel' spin relaxation in a quantum wire with a single subband occupied (recall Problem 8.2). Therefore, there is virtually no spin relaxation in a quantum wire channel if there is no channel magnetic field.

If a channel magnetic field is present, then the spin polarizations in both bands become wavevector dependent and neither subband has a fixed spin quantization axis. Thus, two states in two spin bands, with different wavevectors, have different spin polarizations that are not completely orthogonal. A scattering event (due to a non-magnetic impurity, phonon, etc.) can cause a transition between these states because the matrix element for scattering between them is non-zero. Any such scattering will rotate the spin since the initial and final states have different spin polarizations. Therefore, ordinary non-magnetic scatterers can *relax spin* and the Elliott-Yafet mechanism is revived in a quantum wire channel [7]. Any such spin relaxation will randomize the spin polarizations of electrons arriving at the drain and cause additional leakage current.

There is a way to reduce (or nearly eliminate) the leakage current due to the channel magnetic field. In addition to the Rashba spin-orbit interaction, there is also the Dresselhaus spin-orbit interaction in any semiconductor that lacks crystalline inversion symmetry. The ideal SPINFET analysis of Section 12.2 ignored this interaction, but it is there. In a one-dimensional SPINFET, the pseudo magnetic field due to the Dresselhaus interaction $B_{Dresselhaus}$ is directed along the channel and since it depends on the carrier velocity v_x (see Equation (8.18)), it can be altered by altering the average velocity of carriers in the channel. Therefore, it can be manipulated to partially offset the channel magnetic field caused by the ferromagnetic contacts. This can reduce the leakage current as was shown in ref. [8].

The total magnetic field that an electron will see in a quantum wire SPINFET is

$$\vec{B}_{total} = B_{Rashba}\hat{z} + B_{channel}\hat{x} + B_{Dresselhaus}\hat{x}. \tag{12.8}$$

The last two terms are magnetic fields directed along the channel, in the x-direction. These two fields will cancel each other if

$$B_{Dresselhaus} = \frac{2m^* a_{42}}{g\mu_B \hbar}\left[\left(\frac{m\pi}{W_z}\right)^2 - \left(\frac{n\pi}{W_y}\right)^2\right] v_x = -B_{channel}\ , \tag{12.9}$$

where a_{42} is a material constant, W_y, W_z are the transverse dimensions of the channel, and v_x is the carrier velocity. It is obvious that perfect cancellation is possible only for the carrier velocity v_x satisfying the condition

$$v_x = \left| \frac{B_{channel} g \mu_B \hbar}{2m^* a_{42} \left[\left(\frac{m\pi}{W_z} \right)^2 - \left(\frac{n\pi}{W_y} \right)^2 \right]} \right| . \tag{12.10}$$

Since different carriers will have different velocities at any non-zero temperature and in the presence of scattering, we can never completely eliminate leakage current by playing off the Dresselhaus interaction against the channel magnetic field. However, with a back gate terminal, we can adjust the channel carrier concentration so that the *average velocity* of the carriers satisfies the above equation. That will result in the minimum leakage current.

Structural modification to reduce the channel magnetic field: It should be obvious to the reader that the SPINFET works just as well if the contacts are magnetized in the +y direction instead of the +x direction. Electrons will enter the channel from the source with their spins polarized in the +y direction, and these spins will precess in the x-y plane when the gate voltage is turned on. Therefore, there is no difference between having the contacts magnetized in the y-direction versus having them magnetized in the x-direction, as long as the gate electric field is in the y-direction and the current flow is in the x-direction so that the pseudo magnetic field associated with the Rashba interaction is in the z-direction. The advantage of having the contacts magnetized in the y-direction is that the channel magnetic field will be much weaker now since it will be only a fringing field. This reduces leakage current.

12.2.3 The quantum well SPINFET

The original SPINFET proposed in ref. [2] was not a quantum wire SPINFET that we have described here, but a quantum well (two-dimensional) SPINFET. For the quantum well SPINFET, the Hamiltonian describing an electron in the channel is

$$\begin{aligned} H &= H_0 + H_R = \frac{\hbar^2}{2m^*} \left(k_x^2 + k_z^2 \right) + \epsilon_n + \eta \left(\sigma_z k_x - \sigma_x k_z \right) \\ &= \begin{pmatrix} \frac{\hbar^2}{2m^*} \left(k_x^2 + k_z^2 \right) + \epsilon_n + \eta k_x & -\eta k_z \\ -\eta k_z & \frac{\hbar^2}{2m^*} \left(k_x^2 + k_z^2 \right) + \epsilon_n - \eta k_x \end{pmatrix} , \end{aligned} \tag{12.11}$$

where we assume that the quantum well is in the x-z plane and ϵ_n are the transverse subband bottom energies due to confinement in the y-direction. The reader can easily deduce this Hamiltonian from the material presented in Chapter 7.

If $k_z = 0$, then the eigenspinors of the above Hamiltonian are $[1 \quad 0]^\dagger$ and $[0 \quad 1]^\dagger$, which are the +z-polarized and -z-polarized spins. This means that B_{Rashba} is directed along the z-axis. But if $k_x = 0$, the eigenspinors are $(1/\sqrt{2})[1 \quad 1]^\dagger$ and $(1/\sqrt{2})[1 \quad -1]^\dagger$ which are the +x-polarized and -x-polarized spins. In that case, B_{Rashba} is directed along the x-axis. The pseudo magnetic field due to the Rashba interaction is always directed in the direction that is mutually perpendicular to the gate electric field and the direction of carrier velocity.

If the source injects x-polarized spins, then electrons which have no component of momentum along the channel ($k_x = 0$), do not undergo any spin precession at all since the original spin polarization and the pseudo magnetic field are both in the same direction (x-direction). In this case

$$\frac{d\vec{S}}{dt} = \vec{\Omega} \times \vec{S} = \frac{g\mu_B \vec{B}_{Rashba}}{\hbar} \times \vec{S} = 0 \tag{12.12}$$

so that spin does not precess. But this does not matter so much since these electrons have no motion along the channel and therefore do not contribute to current, in any case. On the other hand, electrons that have some component of momentum along the channel ($k_x \neq 0$), do undergo spin precession. However, the axis of spin precession depends on the ratio of k_x and k_z. This has a deleterious effect when one ensemble averages of k_x and k_y. The net effect is to give rise to a larger leakage current in the off-state.

This effect was recognized in ref. [2] which advocated the use of a narrow channel that approximates a quantum wire channel. A quantum wire SPINFET is superior to a quantum well SPINFET in producing lower leakage current, but on the flip side, the maximum conductance of a true quantum wire channel is the Landauer conductance e^2/h per spin. The maximum conductance of a quantum well channel, on the other hand, is Ne^2/h per spin, where N is the number of transverse subbands along the z-direction that are occupied by carriers. The number N can be quite large for a quantum well. For a channel width of 1 μm and a Fermi De Broglie wavelength of 10 nm, N = 200. Therefore, a quantum well channel has a larger channel conductance which can ultimately lead to a larger device transconductance. The transconductance determines important figures of merit, such as transistor gain and bandwidth. Therefore, a quantum well channel may be preferred over a quantum wire channel in such applications (mostly analog applications) where the ratio of on to off conductance is not the main consideration, but the transistor gain and bandwidth are more important.

12.2.4 The SPINFET based on the Dresselhaus spin-orbit interaction

The SPINFET of ref. [2] was based on the Rashba interaction. The Dresselhaus spin-orbit interaction can also be gainfully employed to realize a different kind of SPINFET [9]. In a one dimensional channel with crystallographic

inversion asymmetry, the strength of the pseudo magnetic field due to the Dresselhaus interaction depends on the physical width of the channel W_z (see Equation (12.9). If we define the one-dimensional channel by a split-gate, then we can vary the potential applied on the split-gate to change the channel width W_z and therefore the strength of the pseudo magnetic field $B_{Dresselhaus}$ associated with the Dresselhaus interaction. This field also causes spins to precess in space at a rate independent of the carrier velocity[§]. By varying the split-gate voltage, one can change $B_{Dresselhaus}$ and therefore the precession rate. As a result, the gate voltage modulates the angle by which the spins precess as they traverse the channel from source to drain. This, in turn, modulates the source-to-drain current. The source and drains are magnetized either in the y-direction or z-direction since the Dresselhaus field $B_{Dresselhaus}$ is in the x-direction if we have the channel in the [100] crystallographic direction. Since the split-gate voltage can be used to modulate the source-to-drain current, we can realize transistor action. The schematic of a Dresselhaus type SPINFET is shown in Figure 12.5. There is no magnetic field along the channel (except weak fringing fields) since the ferromagnetic contacts are magnetized in the y-z plane, which is transverse to the axis of the channel. This reduces the leakage current. The SPINFET based on Dresselhaus interaction has been discussed in ref. [9]. Here, we present the basic ingredients.

Consider a one-dimensional channel along the [100] crystallographic direction that has Dresselhaus interaction but no Rashba interaction and no magnetic field. This channel is defined by split-gates on a two-dimensional electron gas in the x-z plane, so that the confining potential in the z-direction is parabolic. We will assume that the curvature of this parabolic confining potential is ω, so that the effective width of the channel in the z-direction is $\sqrt{\hbar/(2m^*\omega)}$. Current flows in the x-direction. We will make the same assumptions that were implicit in the Datta-Das SPINFET, namely that the source injects spin with 100% efficiency, the drain detects them with 100% efficiency and finally there is no spin relaxation in the channel which is much shorter than the spin relaxation length.

The Hamiltonian describing this channel is

$$H = \epsilon + \frac{\hbar^2 k_x^2}{2m^*} + 2a_{42}\sigma_x k_x \left[\frac{m^*\omega}{2\hbar} - \left(\frac{\pi}{W_y}\right)^2\right], \tag{12.13}$$

where ϵ is the lowest subband energy in the channel and W_y is the width of the two-dimensional electron gas (in the y-direction). We are assuming that only the lowest subband is filled with carriers.

Diagonalization of the Hamiltonian in Equation (12.13) shows that the eigenspinors in the channel are $(1/\sqrt{2})[1\ 1]^\dagger$ and $(1/\sqrt{2})[1\ -1]^\dagger$ which are

[§]The velocity independence should be obvious since the pseudo magnetic field due to the Dresselhaus interaction is proportional to the carrier velocity (see Equation (12.9), just like the pseudo magnetic field due to the Rashba interaction.

+x-polarized and -x-polarized states. They have eigenenergies that differ by $2\beta k_x$ where $\beta = 2a_{42}[m^*\omega/(2\hbar) - (\pi/W_y)^2]$. Therefore, the energies of +x polarized and -x-polarized states will be

$$E(+x-pol) = \epsilon + \frac{\hbar^2 k_x^2}{2m^*} + \beta k_x,$$
$$E(-x-pol) = \epsilon + \frac{\hbar^2 k_x^2}{2m^*} - \beta k_x. \tag{12.14}$$

An electron entering the channel with an energy E will have two different wavevectors k_{x+} and k_{x-} for +x-polarized and -x-polarized spins. If we inject only +z-polarized spins from the ferromagnetic source contact (magnetized in the +z direction), then the incident electron will transmit equally into +x and -x polarized states since

$$\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \frac{1}{\sqrt{2}}\frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 1 \end{bmatrix} + \frac{1}{\sqrt{2}}\frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ -1 \end{bmatrix}. \tag{12.15}$$

The eigenspinor describing the electron's spin at the drain end will be

$$[\Psi] = \frac{1}{2}\begin{bmatrix} e^{ik_{x+}L} + e^{ik_{x-}L} \\ e^{ik_{x+}L} - e^{ik_{x-}L} \end{bmatrix}, \tag{12.16}$$

where L is the channel length.

If the drain is also magnetized in the +z direction and is an ideal half metallic ferromagnet that transmits only the majority spin (i.e., an ideal spin detector), then the transmission probability of the electron (and hence the linear response source-to-drain conductance G_{SD}) will be given by

$$\begin{aligned} G_{SD} &\propto \left| \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} e^{ik_{x+}L} + e^{ik_{x-}L} \\ e^{ik_{x+}L} - e^{ik_{x-}L} \end{bmatrix} \right|^2 \\ &\propto cos^2\left(\frac{(k_{x+} - k_{x-})L}{2}\right) \\ &= cos^2\left(m^*\beta L/\hbar^2\right) \\ &= cos^2\left(\frac{\Phi}{2}\right), \end{aligned} \tag{12.17}$$

where we have used Equation (12.14) to come up with the last two equalities. The angle

$$\Phi = \frac{2m^*\beta L}{\hbar^2} = \frac{4m^* a_{42} L}{\hbar^2}\left[\frac{m^*\omega}{2\hbar} - \left(\frac{\pi}{W_y}\right)^2\right]. \tag{12.18}$$

We can modulate the curvature of the confining potential ω in the z-direction with split-gate potentials which will change Φ and therefore the source-to-drain conductance G_{SD}. This realizes transistor action.

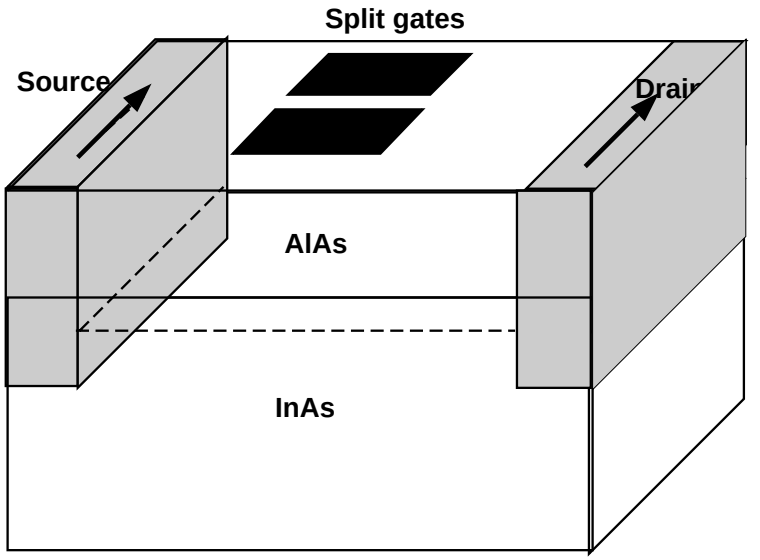


FIGURE 12.5
Schematic of a Spin Field Effect Transistor based on the Dresselhaus interaction. The 2-DEG is shown as the sheet enclosed by the broken lines.

In order to precess the spin by π radians and switch the transistor from "on" to "off", we need

$$\Delta\Phi = \frac{4m^* a_{42} L}{\hbar^2}\left[\frac{m^*\Delta\omega}{2\hbar}\right] = \pi, \tag{12.19}$$

which mandates that

$$\Delta\omega = \frac{\pi\hbar^3}{2m^{*2}a_{42}L}, \tag{12.20}$$

where $\Delta\omega$ is the amount by which we have to change the curvature of the parabolic confining potential in the z-direction with the split-gate voltage.

Just as in the case of the Datta-Das SPINFET, we will benefit from a stronger spin-orbit interaction strength a_{42} and a long channel length L, since these would reduce the split-gate voltage required to turn the transistor from on to off.

12.3 Device performance of SPINFETs

In the world of electronics, the universally accepted benchmark for the transistor device is the celebrated metal-insulator-semiconductor-field effect-transistor (MISFET) which has been, and remains, the workhorse of electronics. Therefore, the SPINFET must be compared with an equivalent MISFET to determine if there are any advantages to utilizing spin. Surprisingly, in spite of many papers extolling perceived merits of SPINFETs, this elementary exercise was not carried out until recently [10]. The results were quite illuminating.

Ref. [10] compared a strictly one-dimensional (quantum wire) MISFET with a strictly one-dimensional SPINFET at very low temperatures and showed that the SPINFET is actually *inferior* (requires a larger gate voltage to switch states) unless its channel is several micrometers long. Here, we extend that analysis to quasi two-dimensional channel with narrow width and at arbitrary temperatures.

We will compare primarily the switching voltage of a two-dimensional MISFET with that of a two-dimensional SPINFET (i.e., the voltage required to turn the device on or off) since that determines the dynamic energy dissipated during switching. The dissipated energy is $(1/2)C_g V_{switch}^2$ where C_g is the gate capacitance[¶]. Therefore a lower V_{switch} translates to less energy dissipated during switching. This can be a significant advantage in high density circuits where the primary obstacle to continuing device miniaturization is excessive energy dissipated during switching.

Let us focus on a narrow GaAs channel SPINFET where the width of the channel is sufficiently small that the component of an electron's velocity along the width is much smaller than the component along the length ($v_x >> v_z$ or $k_x >> k_z$). In that case, Equation (12.5) approximately holds and the electric field $\mathcal{E}_y$ required to turn off the SPINFET is found by setting $\Phi = \pi$ and solving for $\mathcal{E}_y$. This value is

$$\mathcal{E}_y^{off} = \frac{\pi \hbar^2}{2m^* a_{46} L}. \tag{12.21}$$

In GaAs, the value of $a_{46} \approx 10^{-38}$ C-m^2 and m^* is 0.067 times the free electron mass. Therefore, if the channel length $L = 1$ μm, then $\mathcal{E}_y^{off} = 280$ kV/cm. For a 0.1 μm long channel, $\mathcal{E}_y^{off} = 2.8$ MV/cm, which is beginning to approach the breakdown field in GaAs. The field would have been smaller if the value of a_{46} were larger. Unfortunately, there is no known material where the value of a_{46} is significantly (a few orders of magnitude) higher than

[¶]The energy dissipated can be much smaller than $(1/2)C_g V_{switch}^2$ if the gate voltage is applied in small steps. This mode of switching, known as "adiabatic switching", has been discussed by some authors [11], but is slow and unreliable.

what it is in GaAs. Therefore, the practicality of deep sub-micron n-channel SPINFETs is currently dubious since the value of $\mathcal{E}_y^{off}$ in such short channel SPINFETs may exceed the breakdown field in the channel.

There is however a small caveat. It is well known that spin-orbit interactions in many semiconductors are much stronger in the valence band than in the conduction band. This, of course, immediately brings up the possibility that p-channel SPINFETs might require a smaller electric field to turn off, and therefore may produce better SPINFETs. However, spin transport in the valence band is complicated and it is not really clear whether there will be any significant advantage with p-channel devices. This remains an issue for further research.

12.3.1 Comparison between MISFET and SPINFET

The SPINFET, as proposed in ref. [2], is a normally-on device. Therefore, it is best implemented with a standard MISFET (with ferromagnetic contacts) in "accumulation mode", where a conducting channel is already formed when the gate voltage is zero. In standard MOS theory based on the so-called delta depletion approximation [12], the gate voltage V_G applied on the gate of a MISFET that is in accumulation will be dropped entirely across that gate insulator, with practically no drop on the semiconductor channel, since the accumulation layer charge will shield the semiconductor underneath from the electric field. Therefore,

$$V_G \approx \mathcal{E}_i d \ , \tag{12.22}$$

where $\mathcal{E}_i$ is the electric field in the gate insulator (assumed uniform since there are no fixed charges in an ideal gate insulator). Next, Poisson's equation dictates that

$$\kappa_s \mathcal{E}_y = \kappa_i \mathcal{E}_i \ , \tag{12.23}$$

where κ_s and κ_i are the dielectric constants of the semiconductor and insulator, respectively, and $\mathcal{E}_y$ is the transverse electric field in the semiconductor channel at the interface with the gate insulator. This field induces the Rashba interaction. Here, we have assumed that there is no sheet charge at the insulator-semiconductor interface due to such things as interface states.

Combining the two preceding equations, we obtain

$$V_G = \frac{\kappa_s}{\kappa_i} \mathcal{E}_y d. \tag{12.24}$$

Therefore, using Equation (12.21) in Equation (12.24), we find that the gate voltage required to turn off a SPINFET will be

$$V_{off}^{SPINFET} = V_{switch} = \frac{\kappa_s}{\kappa_i} \mathcal{E}_y^{off} d = \frac{\kappa_s}{\kappa_i} \frac{\pi \hbar^2}{2m^* a_{46}} \frac{d}{L}. \tag{12.25}$$

If we assume that the gate insulator is AlGaAs, then $\kappa_s \approx \kappa_i$. Therefore, if the gate insulator thickness is 10 nm, then for a 1 μm channel length device,

$V_{off}^{SPINFET} = 284$ mV. For a 0.1 μm channel length device, $V_{off}^{SPINFET} = 2.84$ V. Remember that in the 0.1 μm device, the transverse electric field required to turn the device off is 2.8 MV/cm, which is close to the breakdown electric field in GaAs. Therefore, GaAs SPINFETs with channel lengths much smaller than 0.1 μm are not very practical. One should ruminate over this, knowing that silicon MOSFETs with gate lengths of 0.09 μm are currently in production, and 0.01 μm gate length silicon MOSFETs have been demonstrated.

Now, if the same SPINFET structure were used as a traditional normally-on MISFET, then what would have been the switching voltage that would have turned the device off? We would simply have to put enough negative charges on the gate to drive all the electrons out of the channel in order to turn the MISFET off. That means

$$C_g V_{off}^{MISFET} = -en_s A \ , \tag{12.26}$$

where e is the electronic charge, n_s is the average surface concentration of electrons in the channel and A is the cross sectional area of the gate (which, we assume, covers the entire channel). Since $C_g = \kappa_i A/d$, we get

$$V_{off}^{MISFET} = V_{switch} = -\frac{en_s d}{\kappa_i}. \tag{12.27}$$

If $n_s = 10^{11}$ cm^{-2} (typically obtained in GaAs channels), and $d = 10$ nm, then $V_{off}^{MISFET} = 14$ mV, independent of the channel length. From Equation (12.25), we see that this is also the $V_{off}^{SPINFET}$ of a SPINFET with 20 μm channel length. Therefore, if a typical GaAs accumulation mode MISFET structure is used either in the SPINFET mode or in the traditional MISFET mode, the SPINFET mode will have a lower switching voltage if the channel length is larger than 20 μm. Consequently, the SPINFET beats the MISFET only if its channel length is huge – larger than 20 μm! This comparison is valid at any arbitrary temperature.

If we use different material systems (e.g., InAs/AlAs heterostructures instead of AlGaAs/GaAs heterostructures), then the numbers will change slightly, perhaps in favor of the SPINFET, but the basic conclusion will not change. When all is said and done, the spin-orbit interaction in technologically important semiconductors is too weak to produce sufficient spin precession rate (in space) at small enough gate voltages. Therefore, not enough spin precession is produced to rotate the spin by π radians if the channel length is short and the gate voltage is not too high. In order to produce a rotation by π, one needs either very high gate voltages or very long channels. That is why the switching voltage of a SPINFET is typically larger than that of a comparable MISFET unless the channel of the SPINFET is unnaturally long. This conclusion is consistent with that reached in ref. [10].

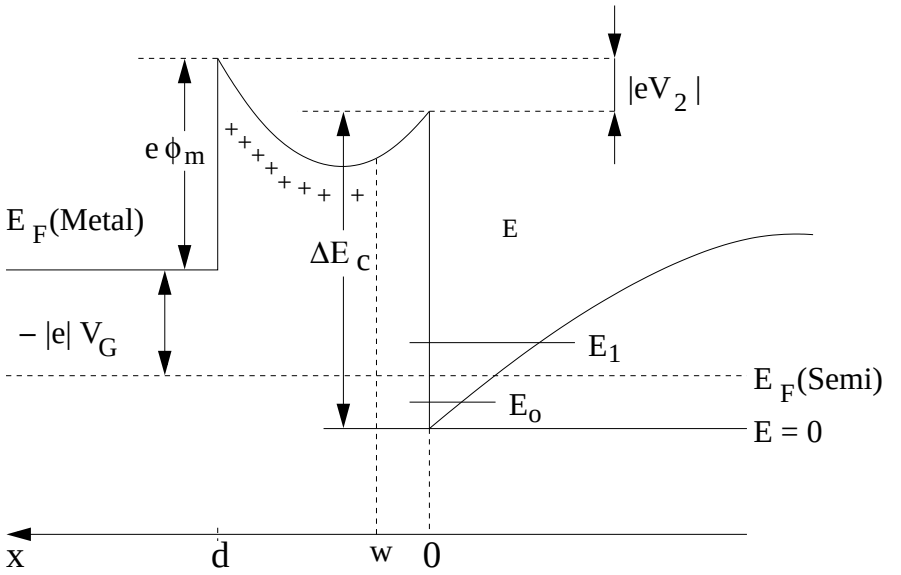


FIGURE 12.6
Energy band diagram in the direction perpendicular to the heterointerface in a HEMT, when a gate voltage V_G is applied. The conduction band offset between the gate insulator material and the semiconductor substrate material is ΔE_c and ϕ_m is the metal-insulator barrier. The Fermi level is denoted by E_F. The conduction band edge in the semiconductor substrate, adjoining the interface, has a triangular shape (see Chapter 15 for a discussion of triangular potential wells). The lowest two bound state energies in the triangular well are E_0 and E_1.

12.3.2 Comparison between HEMT and SPINFET

If a high electron mobility transistor (HEMT) [also known as a modulation doped field effect transistor, or MODFET] is used instead of a MISFET structure to implement a SPINFET, then the analysis is slightly different. The HEMT is a normally-on device, just like an accumulation mode MISFET. Here, however, the gate insulator is doped so that there are fixed charges in the insulator. The average electric field in the gate insulator is typically negligible compared to the electric field in the semiconductor right at the interface with the insulator. The energy band diagram of a HEMT as a function of distance in the direction perpendicular to the heterointerface is shown in Fig. 12.6.

Using Poisson equation, and neglecting the electric field in the insulator in comparison with the electric field in the semiconductor at the insulator-

semiconductor interface, we can write

$$\mathcal{E}_y \approx \frac{en_s}{\kappa_s}, \tag{12.28}$$

where n_s is the surface concentration of carriers in the HEMT channel at zero gate voltage. Following ref. [13], we can further stipulate that

$$\begin{aligned} n_s &= C_g(V_G - V_t)/(eA) \\ C_g &= A\kappa_i/(d + \Delta d)\ , \end{aligned} \tag{12.29}$$

where d is the gate insulator thickness, V_t is the threshold voltage (it is a negative quantity) and Δd is the thickness of the two-dimensional electron gas in the channel. Combining the previous two equations, we get

$$\mathcal{E}_y = \frac{\kappa_i}{\kappa_s}\frac{V_G - V_t}{d + \Delta d}. \tag{12.30}$$

The quantity V_t is given by [13]

$$V_t = \phi_m - \Delta E_c - qN_D(d - w)^2/(2\kappa_s) \tag{12.31}$$

and is usually a negative quantity. Here, ϕ_m is the Schottky barrier height between the gate metal and the gate insulator, ΔE_c is the conduction band offset between the insulator and the semiconductor, N_D is the doping density in the insulator and w is the width of the spacer layer separating the doped region of the insulator from the semiconductor channel.

Note from Equation (12.30) that in the HEMT, there is a large transverse electric field $\mathcal{E}_y$ in the channel. Setting this electric field equal to $\mathcal{E}_y^{off}$ given by Equation (12.21), we obtain

$$V'_{off}{}^{SPINFET} = \frac{\kappa_s}{\kappa_i}\frac{\pi\hbar^2}{2m^* a_{46}}\frac{d + \Delta d}{L} - |V_t|. \tag{12.32}$$

Typically $V_t \sim$ -1 V. Therefore, if we assume d = 10 nm, Δd = 8 nm and the channel material is GaAs, then $V'_{off}{}^{SPINFET}$ = 400 mV if the channel length is 1 μm and 4 V if the channel length is 0.1 μm. In comparison, the switching voltage of the HEMT itself is $|V_t|$ = 1 V. Therefore, the SPINFET does not produce any advantage over the HEMT in terms of a lower switching voltage, or lower dynamic power dissipation during switching, if the channel length is smaller than 0.25 μm. Once again, the reason for this is that spin-orbit interactions are too weak to produce sufficient spin precession within a short channel with small gate voltage. In order for a 10 nm channel length SPINFET to have the same switching voltage as a 10 nm channel length HEMT, the strength of Rashba interaction (a_{46}) had to be 25 times larger than what it is.

12.4 Power dissipation estimates

The dynamic energy dissipated during switching a transistor is

$$P_{dynamic} = (1/2)C_g V_{switch}^2. \tag{12.33}$$

We will consider 100-nm channel length SPINFET devices (because 90-nm gate length MOSFETs are currently in production), for which $V_{off}^{SPINFET} =$ 2.8 V if implemented in the MISFET configuration and 4 V if implemented in the HEMT configuration. Assume also that the channel width is 50 nm. In that case, the gate capacitance C_g is 57 aF in the MISFET configuration and 32 aF in the HEMT configuration. Therefore, the dynamic energy dissipated during switching the transistor is 0.22 femto-Joules in the MISFET configuration and 0.26 femto-Joules in the HEMT configuration. If the switching is carried out in $\sim$ 10 ps, then the power dissipated is 22 μW and 26 μW respectively. With a device density of 10^8 transistors/cm^2 (comparable to the Pentium IV chip), the power density will be 22 kW/cm^2 and 26 kW/cm^2, respectively, which are too high (the Pentium IV dissipates less than 100 W/cm^2). Therefore, the SPINFET does not reduce power dissipation unless its channel is at least a few μm long. Having such a large SPINFET is not acceptable since it reduces device density (severalfold from what we currently have in the Pentium IV chip) and goes against the grain of Moore's law that predicts doubling of the device density every 18 months.

12.5 Other types of SPINFETs

Slightly different types of SPINFET ideas have also appeared in the literature. They go by such names as "Non-ballistic SPINFET" [14, 15] or the "Spin Lifetime Transistor" [16].

12.5.1 The non-ballistic SPINFET

This device has a two-dimensional electron gas like an ordinary MISFET and has both Dresselhaus and Rashba spin orbit interactions. The latter can be changed with a gate voltage. Fig. 12.7 shows such a device. Since the channel is two-dimensional, the spin-split bands generally do not have a fixed spin quantization axis even if there is no magnetic field in the channel [meaning that the spin eigenstates are wavevector-dependent; recall Section 7.1.1 with $B_x = B_z = 0$]. The only exception to this situation is when the Rashba and Dresselhaus interactions in the channel have exactly the same

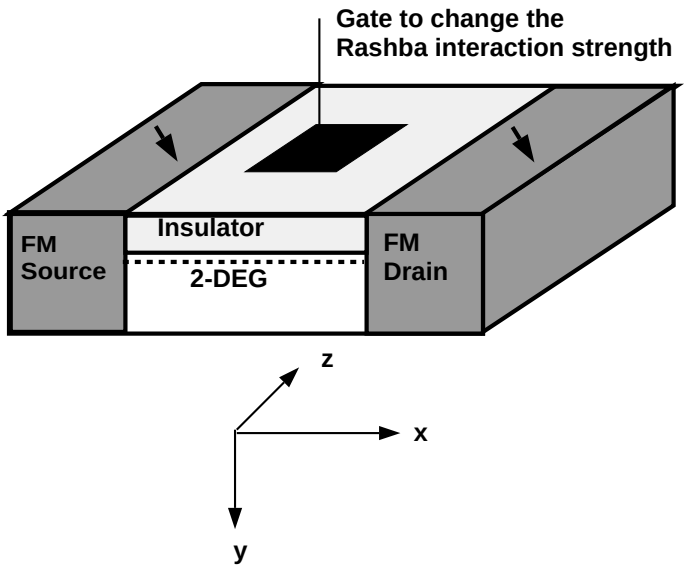


FIGURE 12.7
The non-ballistic SPINFET. The current flows in the x-direction and the symmetry breaking electric field is applied along the y-direction to modulate the strength of the Rashba interaction. When the strengths of the Rashba and Dresselhaus interactions are equal, the wavevector-independent eigenspin polarizations lie in the x-z plane and subtend an angle of 45° or 135° with the x-axis. The contacts are magnetized in the x-z plane subtending an angle of 45° with the x-axis so that they inject and detect spins polarized in one of the two eigenspin states.

strength. In that case, each band has a fixed spin quantization axis, and the spin eigenstate in either band is wavevector-independent. In fact, these eigenspinors are $[-sin(\pi/8), cos(\pi/8)]^T$ and $[cos(\pi/8), sin(\pi/8)]^T$ as can be verified from Equations (7.24) and (7.25). The eigenspin lies in the plane of the 2-DEG (x-z plane) and subtends an angle of 45° or 135° with the x-axis.

In the non-ballistic SPINFET, the Rashba interaction is first tuned with the gate voltage to make it exactly equal to the Dresselhaus interaction which is gate voltage independent. This then makes the spin eigenstates wavevector independent. Electrons are injected into the channel of the transistor from a ferromagnetic source with a polarization that corresponds to the spin eigenstate in one of the spin-split bands. All carriers enter this band. Since the spin eigenstate is wavevector-independent, any momentum relaxing scattering in the channel, which will change the electron's wavevector, will *not alter* the

spin polarization. Scattering can couple two states within the same band, but not in two different bands (with anti-parallel spin quantization axis) since the eigenspinors in these two bands are orthogonal. Therefore, regardless of how much momentum relaxing scattering takes place in the channel, there will be no spin relaxation, and the carriers will arrive at the drain with their spin polarization intact. If the drain ferromagnetic contact is magnetized parallel to the magnetization of the source contact, then all these carriers will exit the device and contribute to current.

In order to change the current, the gate voltage is detuned which makes the Rashba and Dresselhaus interaction strengths unequal, thereby making the spin eigenstates in the channel wavevector dependent (see, again, Equations (7.24) and (7.25)). In that case, if the electrons suffer momentum relaxing collisions due to impurities, defects, surface roughness, etc., their spin polarizations will also rotate resulting in spin relaxation. Thus, the carriers that arrive at the drain no longer have all their spins aligned along the drain's magnetization. Consequently, the overall transmission probability of the electrons decreases, and the current drops. This is how the gate voltage changes the source-to-drain current and produces transistor action.

One simple way of viewing the transistor action is that when the Rashba and Dresselhaus interactions are balanced, the channel current is 100% spin polarized (all carriers arriving at the drain have exactly the same spin, namely the original spin), but when the two interactions are unbalanced, then the spin polarization of the current decreases owing to spin relaxation. The spin polarization can decrease to zero, but *no less*, meaning that on the average, 50% of the spins will be aligned and 50% anti-aligned with the drain's magnetization. The aligned component in the current will transmit and the anti-aligned component will be blocked. Thus, the minimum current (off-current) of this transistor is no less than one-half of the maximum current (on-current). Since the maximum ratio of the on-to-off conductance is only 2, this device is clearly unsuitable for most, if not all, mainstream applications. A recent simulation has shown that the on-to-off conductance ratio is actually only about 1.2 for realistic device parameters [17].

The situation can be improved dramatically if the source and drain contacts have anti-parallel magnetizations, instead of parallel. In that case, when the Rashba and Dresselhaus interactions are balanced, the transmitted current will be exactly zero, but when they are unbalanced, it is non-zero. Therefore, the on-to-off conductance becomes infinity. However, there is a caveat. This argument pre-supposes that the ferromagnetic contacts can inject and detect spins with 100% efficiency, meaning that only the majority spins are injected and transmitted by the ferromagnetic source and drain contacts respectively. If, on the other hand, the spin injection efficiency is only about 70% [4] (which is the record at room temperature at the time of writing this book), so that

$$\frac{I_{maj} - I_{min}}{I_{maj} + I_{min}} = 0.7 \ , \tag{12.34}$$

then 15% of the injected current is due to minority spins. These minority spins will transmit through the drain even when the Rashba and Dresselhaus interactions are balanced. Thus the off-current is 15% of the total injected current ($I_{maj} + I_{min}$), whereas the on-current is still at best 50% of the total injected current. Therefore, the on-to-off ratio of the conductance is 0.5/0.15 = 3.3, which is not a whole lot better than 2. Consequently, achieving a large conductance ratio is very difficult, particularly at room temperature when the spin injection efficiency tends to be small. Furthermore, the drain will also not be a perfect spin filter. The reader can easily show that the ratio of on-to-off conductance of this transistor will be

$$\frac{G_{on}}{G_{off}} = \frac{1}{1 - \zeta_S \zeta_D}, \tag{12.35}$$

where ζ_S is the spin injection efficiency at the source and ζ_D is the spin detection efficiency at the drain. Therefore, in order to make the conductance ratio $\sim 10^5$ which is what today's transistors have, the spin injection and detection efficiencies have to be 99.9995%, which appears impractical at this time.

12.5.2 The spin relaxation transistor

A device very similar to the non-ballistic SPINFET was proposed in ref. [16]. We will call it the Spin Relaxation Transistor (SRT) since it realizes transistor action by modulating the spin relaxation rate in the channel with a gate voltage. With zero gate voltage, the spin-orbit interaction in the channel is weak or non-existent, which makes the spin relaxation time very long. When the gate voltage is turned on, the spin-orbit interaction strength increases, which makes the spin relaxation time short.

The source and drain ferromagnets are magnetized in the anti-parallel configuration. The source injects its majority spins in the channel which arrive with their spin polarizations nearly intact at the drain when the gate voltage is zero, because of negligible spin relaxation in the channel. The drain dutifully blocks them since they are minority spins in the drain contact. Accordingly, the source-to-drain current is nearly zero when the gate voltage is zero.

When the gate voltage is turned on, spins flip in the channel, so that a significant fraction (but at most 50%) of the electrons arrive at the drain with their spins flipped. These electrons are transmitted by the drain because their spins are majority spins in the drain. Therefore, the source-to-drain current increases when the gate voltage is turned on. This device is a normally-off device.

Since the drain transmits only the flipped spins, ref. [16] has proposed the following equation for the source-to-drain current I_{SD}:

$$I_{SD} = \frac{Aen_s}{2\tau_{transit}} \left(1 - e^{-\tau_{transit}/T_1}\right) , \tag{12.36}$$

where A is the cross-sectional area of the channel, e is the electronic charge, and n_s is the two-dimensional electron density in the channel. The quantities $\tau_{transit}$ and T_1 are the transit times through the channel and the spin relaxation time. Since for short channel devices $\tau_{transit} << T_1$ always, the above equation simplifies to

$$I_{SD} \approx \frac{Aen_s}{2T_1}. \tag{12.37}$$

Based on the above, ref. [16] states that the current (or conductance) on-off ratio will be $T_1(V_G = 0)/T_1(V_G = V_t)$, where V_t is the 'threshold voltage' that turns the device fully on and causes the maximum source-to-drain current to flow through the channel. Ref. [16] further assumes that $T_1(V_G = 0) = 1$ μsec and $T_1(V_G = V_t) = 10$ psec, so that the conductance on-off ratio is 10^5.

The bane of 'spin injection efficiency': A little reflection will convince the reader that this analysis tacitly assumes that the spin injection efficiency at the source and the spin filtering efficiency at the drain are both 100%. That assumption is actually implicit in Equation (12.36). If we do not make that unrealistic assumption, then of course this device cannot be any better than the non-ballistic SPINFET with anti-parallel magnetizations of the source and drain contacts. In fact, Equation (12.35) applies for this device as well (see Problem 12.3), so that in order to achieve a conductance on-off ratio of 10^5, the required value of ζ_S or ζ_D is 99.9995%, which is not practical in the near term. If the spin injection and detection efficiencies are 70% each, then the maximum conductance on-off ratio of this device is a mere 1.96, which is far cry from the claimed value of 10^5. The purpose of pointing this out is not to be overly critical of the device design, but to highlight the critical role of spin injection and detection efficiencies in all these devices. Without an extremely high spin injection/detection efficiency, these devices have scant chance of being useful for mainstream applications. Therefore, the bulk of research effort in this area should be directed towards improving spin injection and detection efficiencies in these structures.

Ref. [16] has claimed that their SRT device will have a very low value of the threshold voltage V_t, of the order of 0.1 V. The quantity $1/T_1$ is proportional to the electric field $\mathcal{E}_y^2$. Assuming $n_s = 10^{12}$ cm^{-2}, ref. [16] has claimed that $\mathcal{E}_y = 50$ kV/cm is sufficient to reduce T_1 to 10 psec (from 1 μs when $\mathcal{E}_y = 0$) and turn the device on. Using Equation (12.24), we find that the voltage required to turn a GaAs based SRT on will be 0.05 V if the gate insulator thickness d is 10 nm and $\mathcal{E}_y^{off} = 50$ kV/cm (ref. [16] assumed that effectively $d = 20$ nm, which explains the factor of 2 difference). Here, we have assumed, as always, that the SRT is implemented with a MISFET structure in accumulation. If this structure were used in the MISFET mode instead of the SRT mode, then the MISFET would be normally-on and to turn it off would require a gate voltage of 0.14 V (see Equation (12.27)). A high channel carrier concentration favors the SRT and a low channel carrier concentration favors the MISFET. If the channel concentration were 10^{11} cm^{-2}, then the MISFET

would have had a switching voltage of 14 mV, and the SRT would have had switching voltage in excess of 50 mV. However, the disadvantage of having a high carrier concentration in the SRT is that it tends to degrade the spin injection efficiency even further, leading to even lower conductance on-off ratio and higher leakage current in the off-state. Of course, the MISFET will always have a much higher conductance on-off ratio than the SRT unless extremely high spin injection efficiencies are achieved; thus, the SRT is currently not competitive with a MISFET in any case.

HEMT implementation: If a HEMT structure were used to implement the SRT, the device would be normally-on with anti-parallel magnetizations of the source and drain contacts (since there is already a strong built-in transverse channel electric field in a HEMT when the gate voltage is zero). With parallel magnetizations, the device will be normally-off. This device can be switched "off" (in the case of anti-parallel magnetizations) or "on" (in the case of parallel magnetizations) by applying a negative gate voltage to null the built-in electric field. Since this built-in field has a magnitude of $(\kappa_i/\kappa_s)(|V_t|/(d + \Delta d))$ (see Equation (12.30)), it is obvious that the gate voltage we need to apply to null the electric field is V_t. Therefore, both the HEMT and the SRT will have exactly the same switching voltage. The advantage of using the HEMT structure to implement the SRT is that the leakage current in the off-state will be close to zero (even if the spin injection and detection efficiencies are poor) since application of a gate voltage equal to V_t will not only null the electric field in the channel, but it will also deplete the channel of mobile carriers so that there is no carrier left to carry a leakage current. However, the reduction in the leakage current in this case is not due to any spintronic principle, but rather due to the basic MODFET principle, which switches the transistor off by making the channel carrier concentration go to zero. To sum up, the SRT has no advantage over a conventional HEMT in terms of switching voltage, if implemented with a HEMT structure. However, it may have a *disadvantage*. Since at best only 50% of the injected carriers are transmitted in the SRT when it is in the on-state, while 100% of the injected carriers are transmitted in a ballistic HEMT, the on-current of the HEMT will be at least twice that of the SRT. Since their switching voltages are the same, on the average, the transconductance of the HEMT will be roughly twice that of the SRT. Thus, the SRT is not competitive with a traditional HEMT as a transistor, if implemented with a HEMT-like structure.

12.6 The importance of the spin injection efficiency

It should be obvious to the reader that in all of the SPINFETs discussed so far, the spin injection efficiency is critical. Without a very high spin injection

efficiency, approaching 100%, none of the SPINFETs will have a high enough on-to-off conductance ratio to be of much use in anything. Below, we carry out a systematic analysis of how spin injection efficiency affects the transfer characteristics (source-to-drain current versus gate voltage) of SPINFETs, on-off ratio of the conductance, and the maximum transconductance.

Consider the basic SPINFET proposed by Datta and Das [2]. Since the contacts are magnetized along the direction of current flow (say, the +x-direction), majority spins that are injected from the source into the channel are +x-polarized spins with eigenspinor $(1/\sqrt{2})[1 \;\; 1]^\dagger$ and the minority spins are -x-polarized spins with eigenspinors $(1/\sqrt{2})[1 \;\; -1]^\dagger$. When an electron, that had the majority spin in the source, ultimately arrives at the drain, the spinor describing this electron will be

$$\Psi_{maj} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ e^{i\Phi} \end{bmatrix} \; , \tag{12.38}$$

where Φ is the angle (remember that it is velocity independent) by which the spin has precessed in the channel under the influence of the gate potential. This electron is detected by the majority spin band in the drain with efficiency $(1+\zeta_D)/2$, where ζ_D is the spin detection efficiency at the drain. The probability of this detection is

$$P_{maj}^{maj} = \frac{1+\zeta_D}{2} \left| \frac{1}{\sqrt{2}}[1 \;\; 1] \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ e^{i\Phi} \end{bmatrix} \right|^2 = \frac{1+\zeta_D}{2} cos^2 \left(\frac{\Phi}{2} \right). \tag{12.39}$$

The minority spin band also detects this electron with a probability

$$P_{maj}^{min} = \frac{1-\zeta_D}{2} \left| \frac{1}{\sqrt{2}}[1 \;\; -1] \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ e^{i\Phi} \end{bmatrix} \right|^2 = \frac{1-\zeta_D}{2} sin^2 \left(\frac{\Phi}{2} \right). \tag{12.40}$$

But the source injects majority spins with an efficiency $(1+\zeta_S)/2$, where ζ_S is the spin injection efficiency at the source. Consequently, the component of the source-to-drain current due to majority spin injection from the source is

$$\begin{aligned} I_{SD}^{maj} &= \frac{1+\zeta_S}{4} I_0 \left[(1+\zeta_D) cos^2 \left(\frac{\Phi}{2} \right) + (1-\zeta_D) sin^2 \left(\frac{\Phi}{2} \right) \right] \\ &= \frac{1+\zeta_S}{4} I_0 (1+\zeta_D cos\Phi). \end{aligned} \tag{12.41}$$

Similarly, the component of the source-to-drain current due to minority spin injection from the source is

$$\begin{aligned} I_{SD}^{min} &= \frac{1-\zeta_S}{4} I_0 \left[(1-\zeta_D) cos^2 \left(\frac{\Phi}{2} \right) + (1+\zeta_D) sin^2 \left(\frac{\Phi}{2} \right) \right] \\ &= \frac{1-\zeta_S}{4} I_0 (1-\zeta_D cos\Phi). \end{aligned} \tag{12.42}$$

Since the majority and minority spins are orthogonal in the contacts, the total source-to-drain current is

$$I_{SD} = I_{SD}^{maj} + I_{SD}^{min} = \frac{1}{2} I_0 \left(1 + \zeta_S \zeta_D cos\Phi\right). \tag{12.43}$$

Note from the above equation that if either the spin injection or the spin detection efficiency is zero, then there is no current modulation, since the device does not discriminate between spin polarizations. In that case, the source-to-drain current is constantly $I_0/2$ irrespective of the gate voltage. On the other hand, if $\zeta_S = \zeta_D = 100\%$, then the current oscillates between I_0 and 0, with the average value of $I_0/2$.

The on-current corresponds to $\Phi = 0$ and the off-current to $\Phi = \pi$. Therefore, the ratio of on-to-off conductance is

$$\frac{G_{on}}{G_{off}} = \frac{1 + \zeta_S \zeta_D}{1 - \zeta_s \zeta_D}. \tag{12.44}$$

This ratio is infinity if $\zeta_S = \zeta_D = 100\%$, but drops precipitously to only 9.5 if $\zeta_S = \zeta_D = 90\%$. Fig. 12.8 shows how the conductance on-to-off ratio falls off rapidly with decreasing ζ_D or ζ_S.

The transconductance of the transistor also depends on ζ_S and ζ_D. It is given by

$$\begin{aligned} g_m &= \frac{\partial I_{SD}}{\partial V_G} = -I_0 \zeta_S \zeta_D \frac{m^* a_{46}}{\hbar^2} \frac{\kappa_i}{\kappa_s} sin\Phi \times \begin{cases} L/d & (MISFET\ mode) \\ L/(d + \Delta d) & (HEMT\ mode) \end{cases} \\ g_m^{max} &= \frac{\partial I_{SD}}{\partial V_G} = -I_0 \zeta_S \zeta_D \frac{m^* a_{46}}{\hbar^2} \frac{\kappa_i}{\kappa_s} \times \begin{cases} L/d & (MISFET\ mode) \\ L/(d + \Delta d) & (HEMT\ mode). \end{cases} \end{aligned} \tag{12.45}$$

The transconductance is proportional to the product of the spin injection and detection efficiencies.

Spin injection and detection efficiencies are of paramount importance in determining SPINFET performance. Reference [18] has suggested two possible routes to achieving $\sim 100\%$ spin injection efficiency: (i) the use of 100% spin polarized half metallic ferromagnets as spin injectors, and (ii) the use of spin selective barriers. Unfortunately, there are no 100% spin polarized half metals at any temperature above 0 K because of phonons and magnons [6] and even at 0 K, the 100% spin polarization may be destroyed by surfaces and inhomogeneities [6]. On the other hand, the best spin selective barriers are resonant tunneling devices [19] that have spin-resolved resonant levels. When the carrier energy is resonant with one of these levels, only the corresponding spin transits through the barrier. However, at any temperature exceeding 0 K, these levels are thermally broadened and therefore there is always a non-zero probability of both spins transmitting (albeit unequally), which makes the spin injection efficiency considerably less than 100%. There

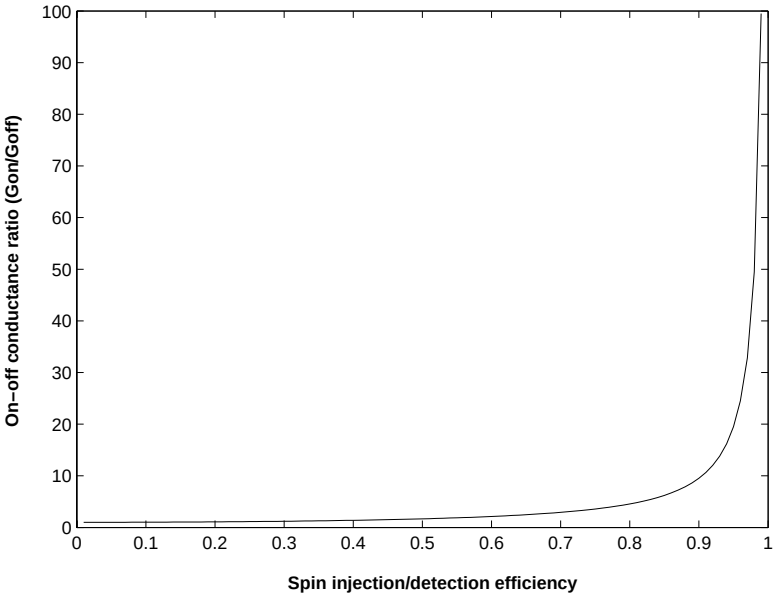


FIGURE 12.8
The ratio of on-to-off conductance of a SPINFET as a function of spin injection and detection efficiency $\zeta = \zeta_S = \zeta_D$.

are other methods for achieving very high spin injection efficiency [20], but they too are ineffective at room temperature. Therefore, it is very difficult to achieve nearly 100% spin injection efficiency at room temperature. As a result, all of the SPINFETs discussed in this chapter are very likely to have either low conductance on-off ratio, or be constrained to work at very low temperatures, or both. Unfortunately, this does not make them competitive with the workhorse of modern electronics - the silicon MISFET.

12.7 Transconductance, gain, bandwidth and isolation

At the beginning of this chapter, we mentioned that a real transistor must have sufficiently high gain and isolation between input and output terminals. The small signal equivalent circuit [21] of the SPINFET is shown in Fig. 12.9. The feedback impedance from the drain node to the gate node is negligible if the gate-to-drain capacitance C_{gd} is small, as it typically is. This indicates that any influence at the output does not back propagate to the input. In other words, the isolation between input and output is good as long as C_{gd} is small.

The transconductance g_m of the device determines the gain and bandwidth. In an accumulation mode transistor, the source-to-drain current does not saturate with increasing source-to-drain voltage. Instead, it continues to increase linearly with voltage (Ohm's law) so that the average source to drain current $I_0/2 = g_0 V_{SD}$. Using this result in Equation (12.45), we obtain

$$\frac{g_m^{max}}{g_0} = \frac{\partial I_{SD}}{\partial V_G} = -2V_{SD}\zeta_S\zeta_D\frac{m^* a_{46}}{\hbar^2}\frac{\kappa_i}{\kappa_s} \times \begin{cases} L/d & (MISFET\ mode) \\ L/(d+\Delta d) & (HEMT\ mode). \end{cases} \tag{12.46}$$

The left hand side of the above equation is the voltage gain of the transistor in the common gate or common source configuration. If we substitute realistic values for the quantities in the right hand side (i.e., use the same values that we used before) and assume 100% spin injection and detection efficiencies, we find that in order to produce a gain larger than unity, V_{SD} must be at least 0.9 V. If $V_{SD} = 5$ V (the normal power supply voltage today), then a GaAs SPINFET's maximum voltage gain is 5.5, if $L = 0.1$ μm and $d = 10$ nm. This is better than that of single electron transistors (SETs) and may be adequate for some applications, but it is certainly not competitive with that of 0.1 μm length Si MOSFETs.

The unity gain frequency (bandwidth) is given by

$$f_T = \frac{1}{2\pi}\frac{g_m}{C_g} \tag{12.47}$$

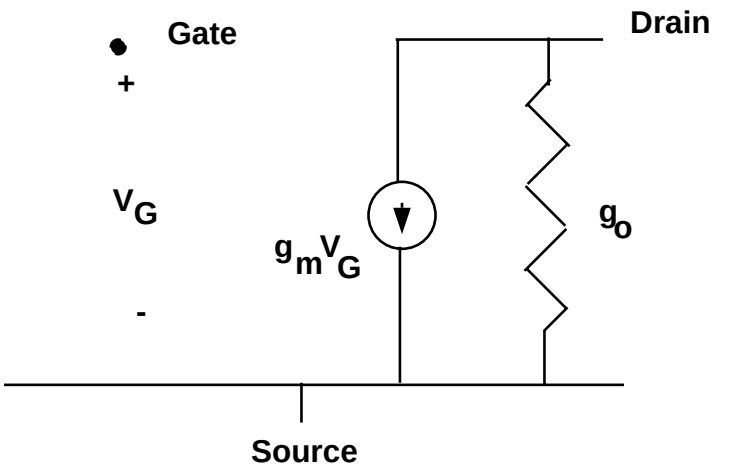


FIGURE 12.9
The small signal equivalent circuit model of a SPINFET.

In a quantum wire SPINFET, the maximum value of g_0 is $2e^2/h = 78$ μSiemens. Therefore, the maximum transconductance of a GaAs SPINFET with a 0.1 μm long channel and a 10 nm thick gate insulator is 430 μSiemens if $V_{SD} =$ 5 V (normal power supply voltage). Since we estimated that $C_g \sim 50$ aF, the unity gain frequency is 10^{13} Hz. This high value, however, is due to the small gate capacitance. It is not a consequence of SPINFET principles.

12.7.1 Silicon SPINFETs

Lately, there has been a spate of activity in silicon based SPINFETs [22, 23] following the successful demonstration of spin injection in silicon [24]. These transistors are a little different from the Datta-Das construct. The Datta-Das device of ref. [2] relies on modulating the strength of Rashba interaction in the channel with a gate voltage to alter spin precession rate and thus realize transistor action. This will not be very effective in silicon which has weak spin-orbit interaction. Therefore Appelbaum and Monsma [23] decided on a different approach. They have an external static magnetic field B_{dc} and spins in the channel precess about this field as they travel from the ferromagnetic source to ferromagnetic drain. The precession rate is the Larmor frequency $g\mu_B B_{dc}/\hbar$. Therefore, the angle by which a spin precesses in traversing the

device is

$$\theta = \frac{g\mu_B B_{dc}\tau_t}{\hbar} = \frac{g\mu_B B_{dc} L}{v\hbar}, \tag{12.48}$$

where $\tau_t = L/v$ is the transit time, v is the electron velocity and L is the distance between the two ferromagnetic contacts.

The angle θ is varied by varying the electron velocity v. That, in turn, is varied by changing the voltage across the channel which changes the accelerating electric field. The drain is a spin selective filter that passes spins aligned parallel to its own magnetization and blocks the anti-parallel spins. Thus, by changing θ with a voltage, the drain current can be modulated and a SPIN-FET is realized. This device has a complicated structure and the interested reader is referred to ref. [23] for a complete description. Huang, et al. [22] have claimed to have demonstrated the operation of this device recently.

One disadvantage of this device is that θ depends on v (unlike in the case of the Datta-Das device) and therefore different electrons with different velocities will precess by different angles. Consequently, ensemble averaging over electrons will dilute the current modulation.

Huang, Monsma and Appelbaum use an interesting approach for spin injection in this device. They use a hot electron emitter that relies on the difference between the scattering rates of majority and minority spins to selectively inject majority spins. For further details of this approach, the reader is referred to refs. [22, 23, 24].

12.8 Spin Bipolar Junction Transistors (SBJT)

The Spin Bipolar Junction Transistor (SBJT) is identical with the normal Bipolar Junction Transistor, except that the base is ferromagnetic and has a non-zero spin polarization. The conduction energy band diagram of a heterojunction npn SBJT is shown in Fig. 12.10. The expressions for the collector (I_C), emitter (I_E) and base (I_B) currents were derived in ref. [25] and are:

$$\begin{aligned} I_C = {} & qA\frac{D_{nb}}{L_{nb}}\frac{1}{sinh(W/L_{nb})}n_{be}\left[e^{qV_{EB}/kT}-1\right] \\ & -qA\frac{D_{nb}}{L_{nb}}coth(W/L_{nb})n_{bc}\left[e^{qV_{CB}/kT}-1\right] \\ & -qA\frac{D_{pc}}{L_{pc}}coth(W_c/L_{pc})p_{oc}\left[e^{qV_{CB}/kT}-1\right] \ , \end{aligned} \tag{12.49}$$

$$I_E = qA\frac{D_{nb}}{L_{nb}}coth(W/L_{nb})n_{be}\left[e^{qV_{EB}/kT}-1\right]$$

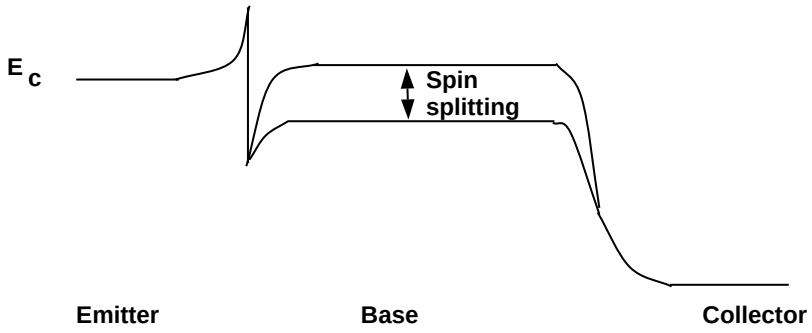


FIGURE 12.10
Energy band diagram showing the emitter, base and collector regions of a SBJT. The base is spin-polarized or ferromagnetic.

$$-qA\frac{D_{nb}}{L_{nb}}\frac{1}{sinh(W/L_{nb})}n_{bc})\left[e^{qV_{CB}/kT}-1\right]$$
$$+qA\frac{D_{pe}}{L_{pe}}coth(W_e/L_{pe})p_{oe}\left[e^{qV_{EB}/kT}-1\right]\ , \qquad (12.50)$$

$$I_E + I_B + I_C = 0, \qquad (12.51)$$

where A is the cross-sectional area of the transistor, W_c is the width of the collector, W_e is the width of the emitter, D_{nb} (L_{nb}) is the minority carrier diffusion constant (length) for electrons in the base, D_{pc} (L_{pc}) is the minority carrier diffusion constant (length) for holes in the collector, D_{pe} (L_{pe}) is the minority carrier diffusion constant (length) for holes in the emitter, $n_{be} = (n_i^2/N_{AB})(1+\alpha_e\alpha_{0b})/\sqrt{1-\alpha_{0b}^2}$, $n_{bc} = (n_i^2/N_{AB})(1+\alpha_c\alpha_{0b})/\sqrt{1-\alpha_{0b}^2}$, $p_{oc} = (n_i^2/N_{DC})$, $p_{oe} = (n_i^2/N_{DE})$, n_i is the intrinsic carrier concentration in the material, N_{AB} is the acceptor dopant concentration in the base, N_{DC} is the donor dopant concentration in the collector, N_{DE} is the donor dopant concentration in the emitter, α_e and α_c are the non-equilibrium spin polarizations in the emitter and collector, α_{0b} ($= tanh(\Delta/kT)$) is the equilibrium spin polarization in the base, and 2Δ is the magnitude of energy splitting between the majority and minority spin in the base.

Based on a small signal analysis, it has been shown that the voltage and current gains of the SBJT are about the same as those of a conventional BJT [26], but the short circuit current gain β ($= \partial I_C/\partial I_B$) has a dependence on the degree of spin polarization in the base (or, equivalently, the spin-splitting energy Δ) which can be altered with an external magnetic field using the

Zeeman effect. Thus, the external magnetic field can act as a *fourth terminal*. This can lead to non-linear circuits such as mixers/modulators. For example, if the ac base current is a sinusoid with a frequency ω_1 and the magnetic field is an ac field which is another sinusoid with frequency ω_2, then the collector current will contain frequency components $\omega_1 \pm \omega_2$. This is one example where spin augments the role of charge, making the SBJT more powerful than the traditional BJT (3-terminal versus 4-terminal).

12.9 GMR-based transistors

The concept of giant magnetoresistance (GMR) was explored in Chapter 11. Its application has been mostly in *passive* devices, such as read-heads. However, it has also spawned a number of proposals for *active* devices (transistors). We will discuss a few of them here.

12.9.1 The all-metal spin transistor

In refs. [27, 28], Johnson proposed and analyzed an all-metal spin-based transistor to capitalize on the first successful reports of GMR measurements in the CPP geometry by Pratt et al. [29] at low temperatures using superconducting electronics and later extended to room temperature by Gijs et al. using microstructure samples [30]. As depicted in Fig. 12.11, Johnson's all-metal spin transistor consists of a spin valve structure with a third terminal connected to the non-magnetic spacer layer. The transistor requires that the thickness of the layers be comparable to or smaller than the spin diffusion length in the material, which is of the order of a few nm in impure metallic samples [31]. An elementary description of the basic principle of operation of this device has been given by Johnson in refs. [27, 28]. Preliminary experimental results were reported by Johnson himself [32], leading only to small voltage output swing and no power or current gain. If such a device could generate a current gain, it could potentially be used to make logic devices. It can also serve as an MRAM. If logic and memory functions can be embedded in the same device, it will yield unprecedented advantages since it obviates the need for communication between logic and memory, which is the primary bottleneck in computation. Moreover, since the electrical characteristics of this purely ohmic device are magnetically tunable, it can potentially be used as a field sensor. Thus, it is a "multifunctional" device with many potential roles that make it interesting.

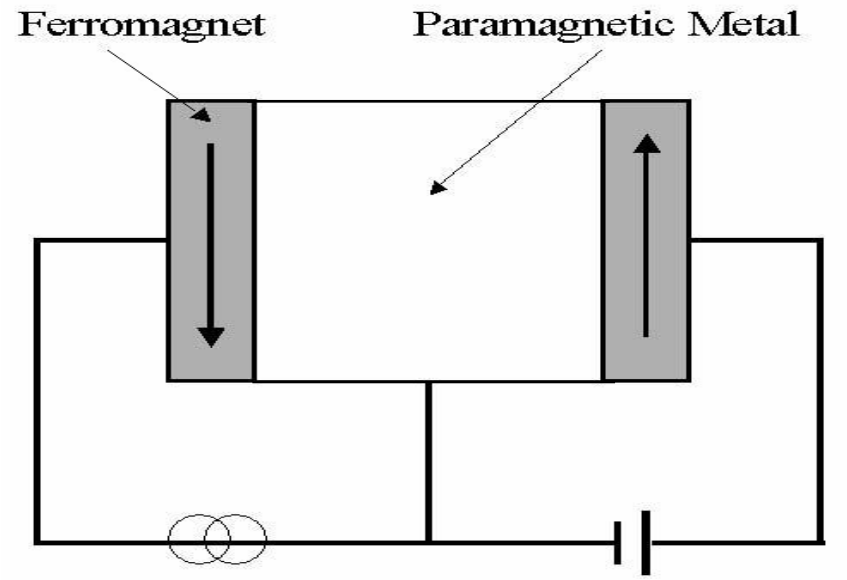


FIGURE 12.11
Schematic of the Johnson spin transistor (after ref. [27]).

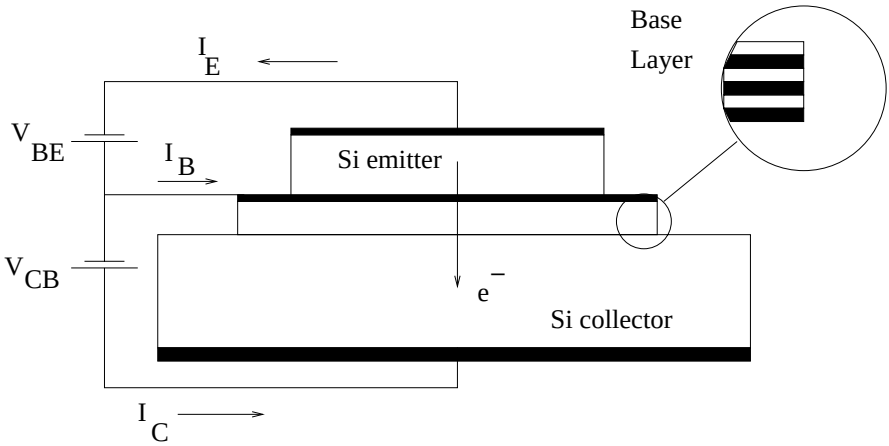


FIGURE 12.12
Schematic cross section of the spin valve transistor proposed by Monsma et al. [33]. The base region is composed of a short superlattice of alternating ferromagnetic and paramagnetic metal layers similar to those used in the demonstration of the giant magnetoresistance effect. The transistor is biased in the common base configuration showing electron injection from the emitter to collector region. Reproduced with permission from D. J. Monsma, et al., Phys. Rev. Lett., **74**, 5260 (1995). Copyright (1995) American Physical Society. http://link.aps.org/abstract/PRL/v74/p5260.

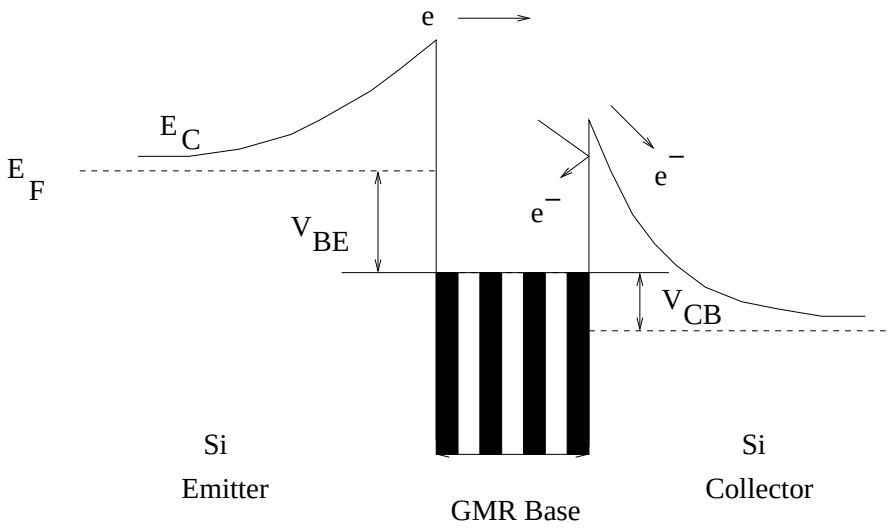


FIGURE 12.13
Schematic energy band diagram from emitter to collector of a spin valve transistor biased in the forward active mode showing the Schottky barriers at the emitter-base and base-collector contacts. Reprinted with permission from D. J. Monsma, et al., Phys. Rev. Lett., **74**, 5260 (1995). Copyright (1995) American Physical Society. http://link.aps.org/abstract/PRL/v74/p5260.

12.9.2 The spin valve transistor

Soon after the proposal by Johnson, Monsma et al. [33] fabricated the first hybrid spintronic device in which ferromagnets and semiconductors were closely integrated (see Fig. 12.12). This device is now usually referred to as the *spin-valve transistor* (SVT) and is basically a reincarnation of the well-known metal base transistor proposed in the 1960s [34]. First fabricated in 1995, the SVT consists of the typical emitter/base/collector configuration of a bipolar junction transistor in which the thin base region is formed of a spin CPP multilayer containing at least two magnetic layers separated by a normal metal spacer. The two magnetic layers act as spin polarizer and analyzer for electrons injected from the n-type emitter. The relative orientation of the magnetic layers in the base region, which can be altered with a magnetic field in the plane parallel to the emitter-base and base-collector interfaces, determines the amount of injected current transmitted across the base and reaching the collector contact. In the active mode of operation, the collector Schottky barrier is reverse-biased and the emitter Schottky is forward-biased (see Fig. 12.13). As a result, in the SVT, the collector current depends strongly on the magnetic configuration of the base region.

In the first SVT report by Monsma et al. [33], a 215% change in the collector current was observed with an applied magnetic field of 500 Oe at

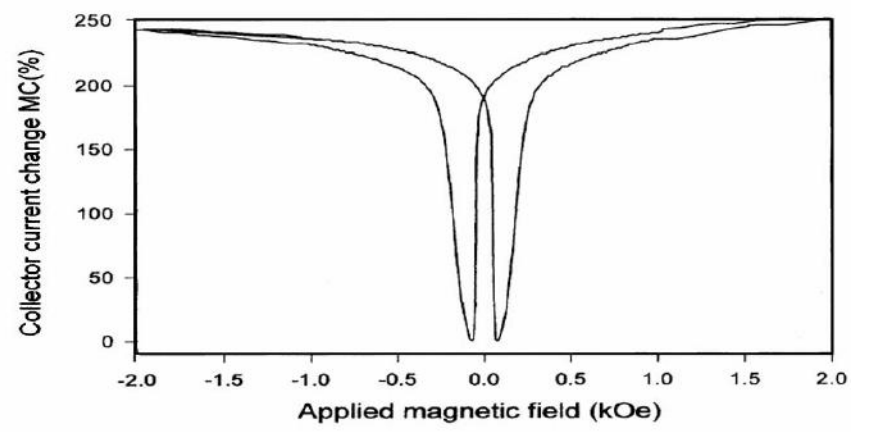


FIGURE 12.14
Magnetocurrent of the first spin valve transistor reported by Monsma et al. [33] with a base composed of four (Cu: 2 nm)/(Co: 1.5nm) bilayers. The measurements were carried out at 77 K. There is a large (390 %) change in the collector current at fairly small value of the external magnetic field. Reprinted with permission from D. J. Monsma, et al., Phys. Rev. Lett., **74**, 5260 (1995). Copyright (1995) American Physical Society. http://link.aps.org/abstract/PRL/v74/p5260.

low temperature (77 K), with the typical hysteresis characteristic of the spin-valve effect (see Fig. 12.14).

As illustrated in Fig. 12.13, the SVT is a hot electron device as a result of the injection of highly energetic (hot) electrons from the emitter into the base. Even if the total base thickness is small (a few hundred angstroms), the hot electrons will lose some energy when crossing the base depending on the bulk (or volume) scattering in each layer and the amount of scattering at the various interfaces bordering the base. This hot electron scattering depends strongly on the magnetic configuration of the multilayers. In the case of anti-ferromagnetic alignment (where magnetizations in adjacent layers are anti-parallel) in the base, both up-spin and down-spin carriers experience heavy inelastic scattering in one of the magnetic layers. As a result, the average kinetic energy of both types of spin injected from the emitter drops exponentially with distance in the base region. On the other hand, for the case of parallel alignment of adjacent ferromagnetic layers in the base, only one type of spins will get scattered heavily whereas the other will reach the base-collector interface without much scattering. In the SVT, the number of hot electrons retaining enough energy to surmount the collector Schottky barrier and contributing to the collector current, is therefore strongly dependent

on the magnetic configuration of the base. In the case of parallel magnetic alignment, more carriers will be emitted over the collector barrier and the collected current will be larger.

The challenge of SVT technology is to design the transistor in such a way that the magneto-current ratio γ_{MC} defined as

$$\gamma_{MC} = (I_P - I_{AP})/I_{AP} \tag{12.52}$$

(where I_P and I_{AP} refer to the collector currents for the parallel (P) and anti-parallel (AP) configurations in the base) is the largest for a given magnetic field applied to switch between the two configurations. Efforts to meet this challenge have led to various modifications of the emitter-base junction design so as to increase the hot electron emitter injection efficiency, and the collector-base junction design to increase the amount of hot carrier collection.

In the original SVT proposed by Monsma et al. [33], the injection energy of hot electrons incident from the emitter is fixed by the height of the Schottky barrier between the emitter and the base. In a related device based on a metal(emitter)/thin insulator/metal(GMR base)/semiconductor(collector) structure proposed by Mizushima et al. [35] (referred to as the magnetic tunnel transistor or MTT), this limitation was overcome by using a tunnel junction as the emitter. The thin insulator acts as a tunnel barrier, injecting hot electrons into the metal base at an energy given by the tunnel bias voltage between the metal emitter and base. This provides additional tunability of the hot electron energy, allowing more detailed spectroscopic studies of spin transport across the base.

One major point of concern in MTT structures is the reliability of the ultra-thin insulator (oxide barrier) which is subjected to a large bias (1-2 V) and large current density. To improve the reliability, LeMinh et al. [36] proposed an MTT device in which the metal emitter is replaced by silicon. As a result, the voltage drop across the emitter-base region is dropped partly across the semiconductor emitter depletion region and partly across the tunnel oxide, thus reducing the voltage stress across the latter. However, the energy of the hot electrons injected into the base is still tunable as the portion of the emitter-base bias dropped across the tunnel insulator raises the conduction band minimum at the semiconductor/tunnel barrier interface with respect to the Fermi level of the base metal. Compared to a magnetic tunnel transistor with a metal emitter, the voltage drop over the thin tunnel oxide is reduced, enabling stable device operation at higher bias voltages. LeMinh et al. recently fabricated such a device with a magneto-current ratio up to 166% and a steep enhanced transfer ratio reaching 6×10^{-4} at an emitter current of 200 mA [36].

In SVT technology, the poor transfer ratio has been the limiting factor leading to current transfer ratio (ratio of the collector current divided by the emitter current) much smaller than unity. This was a major drawback encountered in the early development of the metal base transistor. Despite

this, the Monsma transistor represents a very important step in the evolution of spintronics because it is characterized by electrical characteristics that are magnetically tunable. As pointed out in the recent review article by Jansen on SVT technology [37], the most influential aspect of the Monsma transistor (and its derivatives) is that they showed the value of functional integration of semiconductor and ferromagnetic materials into hybrid electronic devices. Room temperature magneto-current ratio as large as 400% in small magnetic field (typically a few tens of Oe) are now routinely obtained in SVT technology [38].

More recently, the investigation of MTT structures has been successfully extended to GaAs based devices [39]. Much progress has been made in optimizing the device performance, focusing on the output current level and noise sources. Further improvements are still required in order to capitalize on the huge magnetic sensitivity of these structures which make them very attractive candidates for future generations of magnetic sensors or magnetic memories. In that context, MTT structures, scaled to submicron dimensions, will be relevant [37].

SVT and MTT technologies have opened up a new route to the systematic study of the fundamental physics of spin-dependent transport of hot electron at energies of the order of 1 eV. New insights into the fundamental physics of spin-dependent hot-electron scattering have been obtained, including the dominance of the volume effects over interface effects in the spin-dependent transmission across the base, the effects of thermal spin waves, and the surprisingly important role of elastic scattering processes. For instance, it has been realized that hot-electron spin-filtering may have some attractive features for spin-injection into a semiconductor, in particular the ability to reach a spin injection efficiency close to 100 % with conventional ferromagnets. Furthermore, the development of SVT and MTT technologies has led to a detailed study of hot-electron spin-transport through half-metallic ferromagnets and oxides which will lead to the proposal of new types of hybrid electronic devices combining ferromagnets and semiconductors. The status of SVT research was recently reviewed by Jansen [37].

12.10 Concluding remarks

In concluding this chapter, we would like to leave the reader with the impression that hybrid spin devices - such as SPINFETs and SBJTs (including the GMR based SBJTs) - are interesting constructs with interesting physics. They were the pioneers and standard bearers in the field of spintronics, and were primarily responsible for generating interest in this field within the device community. Their performance may not be stellar and they may not

be competitive (currently) with standard semiconductor CMOS or BiCMOS technology for mainstream applications. Nonetheless, interest in these devices persists, not necessarily for killer applications, but more for the wonderful and fascinating physics that they continue to reveal.

12.11 Problems

- **Problem 12.1**

  Consider the one-dimensional Datta-Das SPINFET. Show that the maximum transconductance of this device is

  $$g_m^{max} = \frac{\partial I_{SD}}{\partial V_G} = -\frac{m^* e^2}{2\hbar^3}\frac{\partial \eta}{\partial V_G} L V_{SD}, \tag{12.53}$$

  where I_{SD} is the source-to-drain current, V_G is the gate voltage, V_{SD} is the source-to-drain voltage and L is the channel length.

  In the common-source or common-gate configuration, the voltage gain of the transistor is $g_m/g_o = g_m/G_{SD}$, where G_{SD} is the (linear-response) channel conductance. Show that the maximum voltage gain of the transistor in either of these configurations is

  $$a_v^{max} = \frac{m^*}{\hbar^2} L \frac{\partial \eta}{\partial V_G} V_{SD}. \tag{12.54}$$

  Reference [10] found that for a realistic device configuration, $\partial\eta/\partial V_G \approx 5{\times}10^{-29}$ Coulomb-meter. Using this value and assuming the effective mass of InAs ($m^* = 0.03\ m_0$), show that the voltage gain of a one-dimensional SPINFET (with 100 nm long channel) exceeds unity if V_{SD} exceeds 81 mV. Also, show that for $V_{SD} = 1$ Volt, the maximum voltage gain is greater than 12.

- **Problem 12.2**

  The transconductance g_m of the SBJT will be given by $\partial I_C/\partial V_{BE}$ and the output conductance g_0 will be given by $\partial I_C/\partial V_{CE}$. Using these definitions and the expression for the collector current, show that

  $$g_m \approx \frac{q^2 A}{2kT}\left[\frac{D_{nb}}{L_{nb}}\frac{1}{sinh(W/L_{nb})}\frac{n_i^2}{N_{AB}}(1+\alpha_e tanh(\Delta/kT))\right] e^{q(V_{EB}+\Delta)/kT}$$

  $$g_o \approx \frac{q^2 A}{2kT}\left[\frac{D_{nb}}{L_{nb}} coth\left(\frac{W}{L_{nb}}\right)\frac{n_i^2}{N_{AB}}(1+\alpha_c tanh(\Delta/kT))\right] e^{q(V_{CB}+\Delta)/kT}$$

  $$g_\mu \approx \frac{q^2 A}{2kT}\left[\frac{D_{nb}}{L_{nb}}\frac{W}{2L_{nb}}\frac{n_i^2}{N_{AB}}(1+\alpha_c tanh(\Delta/kT))\right] e^{q(V_{CB}+\Delta)/kT}$$

- **Problem 12.3**

  Consider the ideal SPINFET realized with a quantum wire with a single subband occupied. The axis of the quantum wire is along the x-direction. Show that if B_{Rashba} were proportional to k_x instead of v_x, then the SPINFET will have significant leakage current in the off-state because Φ will be different for different electrons with different velocities. In that case, ensemble averaging will result in a leakage current in the off-state since Φ will never be exactly π radians for every electron in the ensemble. However, since B_{Rashba} is proportional to v_x and not k_x, the leakage current in the off-state is exactly zero since Φ is exactly π radians for every electron in the ensemble regardless of their velocities.

  Hint: Remember Equation (7.53). In the presence of spin-orbit interaction, $v_x \neq \hbar k_x/m^*$.

- **Problem 12.4**

  Show that the conductance on-to-off ratio of the spin relaxation transistor of section 12.5.2 is

$$\frac{G_{on}}{G_{off}} = \frac{1}{1 - \zeta_S \zeta_D}, \tag{12.55}$$

  where ζ_S and ζ_D are the spin injection efficiency at the source/channel interface and spin detection efficiency at the drain/channel interface, respectively.

12.12 References

[1] D. A. Hodges and H. G. Jackson, *Analysis and Design of Digital Integrated Circuits*, 2nd edition, (McGraw Hill, New York, 1988), Chapter 1, p. 2.

[2] S. Datta and B. Das, "Electronic analog of electro-optic modulator", Appl. Phys. Lett., 56, 665 (1990).

[3] E. Y. Tsymbal, O. N. Mryasov and P. R. LeClair, "Spin dependent tunneling in magnetic tunnel junction", J. Phys.: Condens. Matt., **15**, R109 (2003).

[4] G. Salis, R. Wang, X. Jiang, R. M. Shelby, S. S. P. Parkin, S. R. Bank and J. S. Harris, "Temperature independence of the spin injection efficiency of a MgO based tunnel spin injector", Appl. Phys. Lett., **87**, 262503 (2005).

[5] R. Skomski and J. M. D. Coey, *Permanent Magnetism*, (Taylor and Francis, New York, 1999).

[6] P. A. Dowben and R. Skomski, "Are half metallic ferromagnets half metals?", J. Appl. Phys., **95**, 7453 (2004).

[7] M. Cahay and S. Bandyopadhyay, "Phase coherent quantum mechanical spin transport in a weakly disordered quasi one-dimensional channel", Phys. Rev. B, **69**, 045303 (2004).

[8] S. Bandyopadhyay and M. Cahay, "A spin field effect transistor for low leakage current", Physica E, **25**, 399 (2005).

[9] S. Bandyopadhyay and M. Cahay, "Alternate spintronic analog of the electro-optic modulator", Appl. Phys. Lett., **85**, 1814 (2004).

[10] S. Bandyopadhyay and M. Cahay, "Reexamination of some spintronic field effect device concepts", Appl. Phys. Lett., **85**, 1433 (2004).

[11] R. K. Cavin, V. V. Zhirnov, J. A. Hutchby and G. I. Bourianoff, "Energy barriers, demons and minimum energy operation of electronic devices", Fluctuation and Noise Letters, **5**, C29 (2005).

[12] Robert F. Pierret, *Field Effect Devices*, 2nd edition, Modular Series on Solid State Devices, Eds. G. W. Neudeck and R. F. Pierret, (Addison-Wesley, Reading, MA, 1990).

[13] Michael Shur, *Physics of Semiconductor Devices*, (Prentice Hall, Englewood Cliffs, New Jersey, 1990).

[14] J. Schliemann, J. C. Egues and D. Loss, "Non ballistic spin field effect transistor", Phys. Rev. Lett., **90**, 146801 (2003).

[15] X. Cartoixá, D. Z-Y Ting and Y-C Chang, "A resonant spin lifetime transistor", Appl. Phys. Lett., **83**, 1462 (2003).

[16] K. C. Hall and M. E. Flatté, "Performance of spin based insulated gate field effect transistor", Appl. Phys. Lett., **88**, 162503 (2006).

[17] E. Safir, M. Shen and S. Saikin, "Modulation of spin dynamics in the channel of a non-ballistic spin field effect transistor", Phys. Rev. B, **70**, 241302 (2004).

[18] M. E. Flatté and K. C Hall, www.arXiv.org/cond-mat/0604532.

[19] T. Koga, J. Nitta, H. Takayanagi and S. Datta, "Spin filter device based on the Rashba effect using a non-magnetic resonant tunneling diode", Phys. Rev. Lett., **88**, 126601 (2002).

[20] J. Wan, M. Cahay and S. Bandyopadhyay, "Can a non-ideal ferromagnet inject spin into a semiconductor with 100% efficiency without a tunnel barrier?", J. Nanoelectron. Optoelectron., **1**, 60 (2006).

[21] A. S. Sedra and K. C. Smith, *Microelectronic Circuits*, 3rd edition, (Oxford University Press, New York, 1991).

[22] B. Huang, D. J. Monsma and I. Appelbaum, "Experimental realization of a silicon spin field effect transistor", Appl. Phys. Lett., **91**, 072501 (2007).

[23] I. Appelbaum and D. J. Monsma, "Transit time field effect transistor", Appl. Phys. Lett., **90**, 262501 (2007).

[24] I. Appelbaum, B. Huang and D. J. Monsma, "Electronic measurement and control of spin transport in silicon", Nature (London), **447**, 295 (2007).

[25] J. Fabian and I. Zutic, www.arXiv.org/cond-mat/0409196.

[26] S. Bandyopadhyay and M. Cahay, "Are spin junction transistors suitable for signal processing?", Appl. Phys. Lett., **86**, 133502 (2005).

[27] M. Johnson, "Bipolar Spin Switch", Science, **260**, 320 (1993).

[28] M. Johnson, "The all-metal spin transistor", IEEE Spectrum, **31** (5), pp.47-51 (1994).

[29] W.P. Pratt, S.-F. Lee, J.M. Slaughter, R. Loloee, P.A. Schroeder and J. Bass, "Perpendicular giant magnetoresistance of Ag/Co multilayers", Phys. Rev. Lett., **66**, 3060 (1991).

[30] M. A. M. Gijs, S.K.J. Lenczowski and J. B. Giesbers, "Perpendicular giant magnetoresistance of microstructured Fe/Cr magnetic multilayers from 4.2 to 300 K, Phys. Rev. Lett., **70**, 3343 (1993).

[31] I. Zutic, J. Fabian and S. Das Sarma, "Spintronics: Fundamentals and applications", Rev. Mod. Phys., **76**, 351 (2004).

[32] M. Johnson, "Spin polarization of gold films via transport", J. Appl. Phys., **75**, 6714 (1994).

[33] D.J. Monsma, J.C. Lodder, Th. J.A. Popma, and B. Dieny, "Perpendicular hot electron spin valve effect in a new magnetic field sensor: The Spin Valve Transistor", Phys. Rev. Lett., **74**, 5260 (1995).

[34] S.M. Sze, *Physics of Semiconductor Devices*, (Wiley Interscience, New York, 1969), Chapter 11.

[35] K. Mizushima, T. Kinno, T. Yamauchi and K. Tanaka, "Energy dependent hot electron transport across a spin valve", IEEE Trans. Magn., **33** 3500 (1997).

[36] P. LeMinh, H. Gokcan, J.C. Lodder, and R. Jansen, "Magnetic tunnel transistor with a silicon hot electron emitter", J. Appl. Phys., **98**, 076111 (2005).

[37] For a recent review of the basic science and technology of the spin-valve transistor, see R. Jansen, "Spin valve transistor: Review and outlook", J. Phys. D: Appl. Phys., **36**, R289-R308 (2003).

[38] R. Jansen, H. Gokcan, O.M.J. van't Erve, F.M. Postma and J.C. Lodder, "Spin valve transistors with high magnetocurrent and 40 μA output current", J. Appl. Phys., **95**, 6927 (2004).

[39] G. Salis, R. Wang, X. Jiang, R. M. Shelby, S. S. P. Parkin, S.R. Bank and J.S. Harris, "Temperature independence of the spin injection efficiency of a MgO based tunnel spin injector", Appl. Phys. Lett., **87**, 262503 (2005).

13

Monolithic Spintronics: All-spin logic processors

13.1 Monolithic spintronics

In the last chapter, we saw that Spin Transistors - whether they are of the 'field effect' type, or the 'bipolar junction' type - do not really yield significant advantages over their traditional charge based counterparts. Spin transistors are hybrid spin devices where charge still plays the dominant role. Digital information (i.e., binary bit 0 or 1) is ultimately encoded by charge. For example, in the case of the SPINFET, when charge flows through the channel of the transistor causing a source-to-drain current, the device is "on" and could encode the binary bit 1. When no current flows and the device is "off", the encoded bit could be 0. Switching between logic bits is therefore associated with turning on or off a current, which involves *physical motion of charges*. This physical motion consumes considerable energy, which is ultimately dissipated as heat.

Heat dissipation is an extremely serious issue in electronics since it is the primary threat to the fabled "Moore's law". That "law" is an empirical law which predicts that transistor density on a chip will roughly double every 18 months. Moore's law has driven the commercial juggernaut that we call the electronics industry, and it is sacrosanct. Anything that threatens this law is a looming catastrophe which must be eliminated. If heat dissipation cannot be reduced dramatically by switching from electronics to spintronics, and thereby perpetuating Moore's law, then perhaps spintronics has little chance of displacing the silicon juggernaut.

Fortunately, spintronics can, *in principle*, reduce heat dissipation significantly. With charge based electronics, there is always a *fundamental* limitation as far as energy dissipation is concerned. Charge is a *scalar* quantity, which only has a magnitude. Thus, logic levels can be demarcated solely by a difference in the *magnitude* of charge, or by the presence and absence of charge. Consequently, to switch from one bit to another, we must change the magnitude of charge in the active region of the device, or move charge around in space. That invariably causes a current (I) flow and an associated I^2R dissipation (R is the resistance in the path of the current). This dissipation

cannot be avoided.

Spin, unlike charge, is not a scalar quantity. It is a pseudo vector, with a fixed magnitude of $\hbar/2$, but a variable "polarization". We can make the polarization "bistable" by placing an electron in a static magnetic field, so that the polarization has only two allowed states – parallel to the field and anti-parallel. No other polarization is an eigenstate. Spin polarization therefore becomes a *binary* variable. We can encode bits 0 and 1 in these two polarizations. For example, the polarization parallel to the field could encode 1 and the anti-parallel polarization could encode 0, or vice versa. In that case, switching can be accomplished simply by flipping the spin, *without causing physical motion of charges*, or a current flow. This may result in considerable energy saving. There is still some energy dissipated in flipping the spin, but it is of the order of $g\mu_B B$, where B is the flux density of the global magnetic field that we need to keep the spin polarization bistable. This could be made arbitrarily small by making $B \rightarrow 0$. A smaller B of course causes more random bit flips, but bit flip errors can be handled with software or error correcting codes up to a point. In fact, errors occurring with a probability as high as 3% can be handled by the most sophisticated codes that are available today [1]. We will later show that at any temperature T, $g\mu_B B$ must be kept larger than $kTln(1/p)$ where p is the error probability (associated with random bit flips) that we can tolerate or handle with error correcting codes. This is true as long as the spins are in equilibrium with their thermal environment [2]. Therefore, a smaller B results in a higher error probability, since $p > exp[-g\mu_B B/(kT)]$. The point to note here is that we need some minimum energy dissipation $g\mu_B B \geq kTln(1/p)$ *not* because the switching mechanism demands it, but because we have to to keep the error probability manageable.

Since B is required only to make the spin polarization bistable, we ask: why does spin polarization have to be *bistable* in this paradigm? In charge based electronics, logic states are ultimately encoded by voltage or current, which are *continuous* (not discrete) variables. They are certainly not bistable. So, why does spin polarization have to be discrete and bistable? The answer is that in charge based electronics, there are voltage or current amplifiers with *non-linear* transfer characteristics that restore strayed logic levels at circuit nodes to one of two values: 0 and 1. A discussion of this can be found in reference [3]. There is no equivalent device in spintronics to restore logic levels encoded in spin polarization. Therefore, we must ensure that only two spin polarizations are allowed and intermediate polarizations are not stable (or eigenstates). Hence, the need for bistability.

13.1.1 Bit stability and fidelity

In order to encode logic bits reliably, spin polarization must be robust and should not flip (spontaneously) too easily. Spontaneous bit flip is a problem that afflicts charge based electronic memory as well. There, the stored volt-

age level can decay with time owing to charge leakage causing loss of stored data. This is usually countermanded by using a clock to refresh the charge or voltage. The same can be done with spin. We can periodically read the spin polarization of target electrons using a variety of methods. If any spin is found to have flipped spontaneously (as determined by error detecting algorithms), then we can re-write the errant spin by suitable methods. This is equivalent to 'refreshing'.

Of course, we do not want to have to refresh spin bits too often since this can be overwhelming and consumes energy. Therefore, it is imperative that the spin relaxation time is long. If the electron hosting the spin is localized in a quantum dot, then indeed the spin relaxation time can be quite long. In InP quantum dots, the single electron spin flip time has been reported to exceed 100 μs at 2 K [4]. More recently, several experiments have been reported claiming spin flip times (or so-called longitudinal relaxation time T_1) of several milliseconds, culminating in a recent report of 1 second in a GaAs quantum dot at low temperature [5]. Organics are even better because of the weak spin-orbit interaction in these materials. The spin flip time in the π-conjugated molecule tris(8-hydroxyquinolinolato aluminum), popularly known as Alq_3, is reported to possibly exceed 1 second at 100 K [6]. Let us assume conservatively that the spin flip time is 10 msec at the operating temperature. If the clock frequency for refresh cycles is 50 GHz, then the clock period is 20 ps, which is a factor of 5×10^8 smaller than the spin flip time. Therefore, the probability that an unintentional spin flip will occur between two successive refresh clock pulses is $1 - e^{-2\times 10^{-9}} = 2 \times 10^{-9}$, which can be easily handled even in ultra large scale integrated circuits.

13.2 Reading and writing single spin

The next issue to address is how can one align the spin polarization of a single electron in a quantum dot to write data, and how does one read data by measuring a single electron's spin polarization. We can host single electrons in single quantum dots and orient (write) the spin polarization in a target quantum dot by a variety of methods (see, for example, ref. [7]). The easiest method is to generate a local magnetic field with inductors (something routinely done in magnetic random access memory chips) which will orient the local spin parallel to the field and write a bit. Reading is slightly more difficult. There are however three different successful experimental demonstrations of the reading of a single spin in a solid [8, 9, 10]. Among them, the technique of ref. [8] is conceptually the simplest. Here, highly precise magnetic resonance force microscopy is utilized for ascertaining the spin polarization of a single electron in a solid.

13.3 Single Spin Logic

The *Single Spin Logic* (SSL) paradigm [11], proposed in 1994, uses the idea of encoding binary bits in the bistable spin polarization of single electrons housed in quantum dots. A Boolean logic gate, or an entire combinational/sequential logic circuit, can be implemented with a two-dimensional array of single-electron quantum dots with nearest neighbor exchange interaction. Certain dots in the array are designated as input ports and the spins of the single electrons in those dots are oriented and held fixed with some external device. These orientations conform to the input bits. The array is then allowed to relax to the thermodynamic ground state by emitting phonons. Exchange interactions between nearest neighbor quantum dots and the states of the input spins determine the ground state spin polarizations in every dot of the array. By arranging quantum dots in appropriate two-dimensional patterns on a wafer, we can ensure that the spin polarizations in quantum dots that are designated as output spins conform to the desired output(s) in accordance with the truth table of the logic gate or the logic circuit that is being implemented.

Different logic circuits are implemented using different layouts of the quantum dots. We illustrate this idea with the example of a NAND gate. Since this gate is universal, any combinational or sequential logic circuit can be built with the Single Spin NAND gate.

13.3.1 The universal Single Electron Logic gate: the NAND gate

The NAND gate is realized with a linear chain of three electrons in three quantum dots as shown in Fig. 13.1. There is a global dc magnetic field of flux density B to define the spin quantization axis, i.e., spins in any quantum dot will be either parallel or anti-parallel to this field. We will assume that only nearest neighbor electrons interact via exchange since their wavefunctions overlap. Second nearest neighbor interactions are negligible since exchange interaction strength decays exponentially with separation*.

Let us now regard the two outside spins in Fig. 13.1 (dots A and C) as the two inputs to the logic gate and the spin in the middle (dot B) as the output. Assume that the down-spin state ($\downarrow$), which is parallel to the global magnetic field, represents logic bit 1, and the up-spin state ($\uparrow$), which is anti-parallel to the global magnetic field, is logic bit 0. We will show the following result: If we apply logic inputs to the peripheral dots A and C by aligning their spins

*The exchange interaction strength between two electrons depends on the spatial overlap of their wavefunctions. If the electrons are confined in two different quantum dots, the overlap decreases exponentially with increasing separation between the dots.

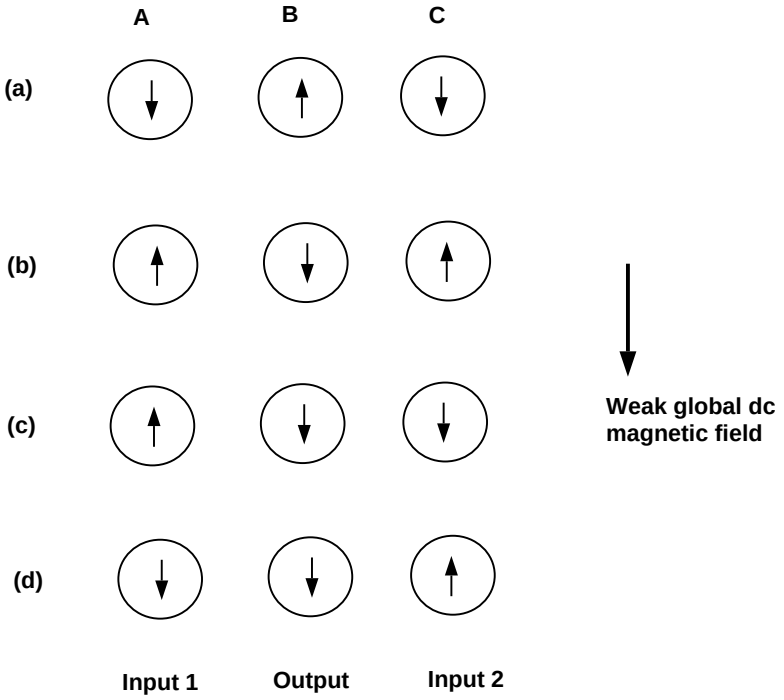


FIGURE 13.1
A 3-dot array with each dot containing a single electron. Nearest neighbors are exchange coupled, i.e., there is exchange coupling between A and B and between B and C, but not between A and C. A global magnetic field is applied to define spin quantization axes. The peripheral dots A and C are the 'input ports' in which the spins are aligned with external agents to conform to input bit 0 or 1. The central dot B is the 'output port' whose spin is the output bit in response to the input bits. Figures (a) - (d) show the ground state configurations when the inputs are aligned as indicated. In all cases, the output bit conforms to the truth table of a NAND gate, showing that this array implements a NAND gate.

either parallel or anti-parallel to the global magnetic field, and then let the system relax to the thermodynamic ground state, the logic output in dot B will always settle to a polarization that will be the correct output conforming to the truth table of a NAND gate:

TABLE 13.1
Truth Table of a Boolean NAND gate

Input 1	Input 2	Output
1	1	0
0	0	1
0	1	1
1	0	1

Once the NAND gate is realized, any arbitrary circuit can be implemented. Sarkar et al. have implemented a large array of combinational and sequential logic circuits using this approach [12]. In the next section, we show that the ground state configurations indeed conform to the rows in Fig. 13.1 which represent the truth table of the NAND gate.

13.3.2 The input dependent ground states of the NAND gate

Consider a linear array of three quantum dots designated A, B and C as in Fig. 13.1. A and C are the input dots and B is the output dot. One can describe this system with the so-called Hubbard Hamiltonian [13]. The Hubbard Hamiltonian is a convenient form of the Hamiltonian for many electron systems. We have not discussed it in this textbook, but excellent descriptions are found in many solid state physics and quantum mechanics textbooks, such as ref. [13]. For our 3-electron system, the Hubbard Hamiltonian is

$$H_{Hubbard} = \sum_{i\sigma} \epsilon_0 n_{i\sigma} + \sum_{input\ \ dots,\ \sigma} g\mu_B B_i sign(\sigma) + \sum_{<ij>} \left[c^{\dagger}_{i\sigma} c_{j\sigma} + h.c. \right] + \sum_{i} U_i n_{i\uparrow} n_{i\downarrow} + \sum_{<ij>\alpha\beta} J_{ij} c^{\dagger}_{i\alpha} c_{i\beta} c^{\dagger}_{j\beta} c_{j\alpha} + B_z \sum_{i\sigma} g\mu_B n_{i\sigma} sign(\sigma) \ , \tag{13.1}$$

where $h.c.$ stands for 'Hermitean conjugate'. Here the first term represents the site energy (kinetic + potential) in the i-th dot, the second term is the Zeeman splitting energy associated with applying a local magnetic field to the i-th dot to selectively align the spin in that dot along the direction of the applied local field (providing input data), the third term is the tunneling between nearest neighbor dots (the angular braket denotes 'nearest neighbors'), the fourth

term is the charging energy of the dot, which is the energy cost associated with putting a second electron in the same dot (Pauli exclusion principle ordains that the second electron must have a spin anti-parallel to that of the first, since we assume that each dot has only one bound state), the fifth term is the exchange interaction between nearest neighbor dots and the last term is the splitting energy due to the global magnetic field, applied over the entire array, in the z-direction, to define the spin quantization axes. Here, we have used the second quantization operators – the so-called creation and annihilation operators. Readers not familiar with these operators can refer to any advanced quantum mechanics textbook where they are discussed. However, for what follows, knowledge and understanding of second quantization operators are not necessary.

If the charging energy (intradot Coulomb repulsion) is sufficiently strong, the Hubbard model at half filling (one electron per dot) reduces to the so called Heisenberg model which was briefly discussed in Chapter 9. This model is much simpler to handle than the full Hubbard model, and yet captures all the essential physics. This was the approach adopted by Molotkov and Nazin in analyzing SSL gates [14, 15]. Using this approach, the Hamiltonian reduces to

$$H_{Heisenberg} = \sum_{<ij>} J_{ij}^{\parallel} \sigma_{zi}\sigma_{zj} + \sum_{<ij>} J_{ij}^{\perp} \left(\sigma_{xi}\sigma_{xj} + \sigma_{yi}\sigma_{yj}\right) + \sum_{input\ dots} \sigma_{zi} h_{zi}^{input} + \sum_{i} \sigma_{zi} h_{zi}^{global}, \tag{13.2}$$

where we have assumed that the input data are provided to the input dots by z-oriented local magnetic fields. We will assume the isotropic case whereby $|J_{ij}^{\parallel}| = |J_{ij}^{\perp}| = J$. Here J is the exchange energy that is non-zero if the wavefunctions in dots i and j overlap in space. Roughly speaking, it is one-half of the energy difference between the triplet state (when spins in dots i and j are parallel) and the singlet state (when the spins in dots i and j are anti-parallel). In the above equation, the σ-s are of course the Pauli spin matrices.

Now, the spins in dots A, B and C can be either +z-polarized (anti-parallel to the global magnetic field) or -z-polarized (parallel to the global magnetic field), which we call "up-spin" and "down-spin", respectively. Remember that the "up-spin" state is logic 0 and "down-spin" state is logic 1. Therefore, the 3-spin basis states that span the entire Hilbert space are $|\downarrow_A\downarrow_B\downarrow_C>$, $|\downarrow_A\downarrow_B\uparrow_C>$, $|\downarrow_A\uparrow_B\downarrow_C>$, $|\downarrow_A\uparrow_B\uparrow_C>$, $|\uparrow_A\downarrow_B\downarrow_C>$, $|\uparrow_A\downarrow_B\uparrow_C>$, $|\uparrow_A\uparrow_B\downarrow_C>$, $|\uparrow_A\uparrow_B\uparrow_C>$. For the sake of brevity, we will write them as $|\downarrow\downarrow\downarrow>$, $|\downarrow\downarrow\uparrow>$, $|\downarrow\uparrow\downarrow>$, $|\downarrow\uparrow\uparrow>$, $|\uparrow\downarrow\downarrow>$, $|\uparrow\downarrow\uparrow>$, $|\uparrow\uparrow\downarrow>$, $|\uparrow\uparrow\uparrow>$. These eight basis functions form a complete orthonormal set. The matrix elements $<\phi_\alpha|H_{Heisenberg}|\phi_\beta>$ are summarized in the matrix below [16]. Here the ϕ-s are the 3-electron basis states enumerated above.

$$\begin{pmatrix} D_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & D_2 & 2J & 0 & 0 & 0 & 0 & 0 \\ 0 & 2J & D_3 & 0 & 2J & 0 & 0 & 0 \\ 0 & 0 & 0 & D_4 & 0 & 2J & 0 & 0 \\ 0 & 0 & 2J & 0 & D_5 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2J & 0 & D_6 & 2J & 0 \\ 0 & 0 & 0 & 0 & 0 & 2J & D_7 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & D_8 \end{pmatrix},$$

where

$$\begin{aligned} D_1 &= 2J - h_A - h_C - 3Z \\ D_2 &= -h_A + h_C - Z \\ D_3 &= -2J - h_A - h_C - Z \\ D_4 &= -h_A + h_C + Z \\ D_5 &= h_A - h_C - Z \\ D_6 &= -2J + h_A + h_C + Z \\ D_7 &= h_A - h_C + Z \\ D_8 &= 2J + h_A + h_C + 3Z. \end{aligned} \tag{13.3}$$

In the above matrix, Z is one-half of the Zeeman splitting energy associated with the global magnetic field, while h_A and h_C are the energies supplied at the inputs. If the local magnetic field providing input to an isolated dot aligns the spin in the direction of the global magnetic field (down-spin direction), then the corresponding h is positive. Otherwise, it is negative. The quantity J in Equation (13.3) is always positive[†].

We will evaluate the 8 eigenenergies E_n ($n = 1....8$) by finding the eigenvalues of the 8×8 matrix above and the corresponding eigenstates

$$\begin{aligned} \psi_n &= c_1^n|\downarrow\downarrow\downarrow> + c_2^n|\downarrow\downarrow\uparrow> + c_3^n|\downarrow\uparrow\downarrow> + c_4^n|\downarrow\uparrow\uparrow> \\ &\quad + c_5^n|\uparrow\downarrow\downarrow> + c_6^n|\uparrow\downarrow\uparrow> + c_7^n|\uparrow\uparrow\downarrow> + c_8^n|\uparrow\uparrow\uparrow> \\ &= [c_1^n, c_2^n, c_3^n, c_4^n, c_5^n, c_6^n, c_7^n, c_8^n]\,. \end{aligned} \tag{13.4}$$

We will repeat this for four cases: $h_A = \pm h$ and $h_C = \pm h$. These four cases correspond to the four entries in the truth table of the NAND gate.

[†]If we evaluate J from the overlap of the wavefunctions in neighboring dots, it will turn out to be negative (see Chapter 9). However, refer back to Equation (9.85), where the Heisenberg Hamiltonian was written as $\widehat{H} = -\frac{1}{2}\sum_{i<j} J_{ij}\,\overrightarrow{\sigma}_i \cdot \overrightarrow{\sigma}_j$. Since we have written the Hamiltonian in Equation (13.2) in the form $+\sum_{i<j} J_{ij}\,\overrightarrow{\sigma}_i \cdot \overrightarrow{\sigma}_j$, our J has to be positive for consistency. A positive J makes the singlet state lower in energy than the triplet state, which is the correct result.

Case I: $h_A = h_C = h > 0$**:** This is the case when input bits [1] and [1] are written in dots A and C by the local magnetic fields. In this case, the eigenenergies and eigenstates are obtained by diagonalizing the matrix in Equation (13.3). They are given in Table 13.2

TABLE 13.2
Eigenenergies and eigenstates when inputs are [1, 1]

Eigenenergy	Eigenstate
$-J-h-Z-\Delta_1$	$[0, 2/\Upsilon_1, \Lambda_1/(J\Upsilon_1), 0, 2/\Upsilon_1, 0, 0, 0]$
$2J-2h-3Z$	$[1, 0, 0, 0, 0, 0, 0, 0]$
$-J+h+Z-\Delta_2$	$[0, 0, 0, 2/\Upsilon_3, 0, -\Lambda_3/(J\Upsilon_3), 2/\Upsilon_3, 0]$
$-Z$	$[0, 1/\sqrt{2}, 0, 0, -1/\sqrt{2}, 0, 0, 0]$
Z	$[0, 0, 0, -1/\sqrt{2}, 0, 0, 1/\sqrt{2}, 0]$
$-J-h-Z+\Delta_1$	$[0, 2/\Upsilon_6, \Lambda_6/(J\Upsilon_6), 0, 2/\Upsilon_6, 0, 0, 0]$
$-J+h+Z+\Delta_2$	$[0, 0, 0, 2/\Upsilon_7, 0, -\Lambda_7/(J\Upsilon_7), 2/\Upsilon_7, 0]$
$2J+2h+3Z$	$[0, 0, 0, 0, 0, 0, 0, 1]$

where

$$\begin{aligned}
\Delta_1 &= \sqrt{(h+J)^2+8J^2},\\
\Delta_2 &= \sqrt{(h-J)^2+8J^2},\\
\Lambda_1 &= -J-h-\Delta_1,\\
\Lambda_3 &= J-h+\Delta_2,\\
\Lambda_6 &= -J-h+\Delta_1,\\
\Lambda_7 &= J-h-\Delta_2\\
\Upsilon_n &= \left[8+(\Lambda_n/J)^2\right]^{1/2}.
\end{aligned} \tag{13.5}$$

In the above table, the eigenenergies are arranged in ascending order (i.e., the first entry is the ground state and the last entry is the highest excited state), *provided* $h >> J$ and $J > Z/2$. The last inequality merely ensures that the first excited state has a higher energy than the ground state. The reason for the first inequality will become clear shortly.

Note that the ground state wavefunction is the entangled state[‡]

[‡]An entangled state is one which cannot be written as the tensor product of constituents. The famous example is the so-called singlet state of two electron spins $(1/\sqrt{2})[|\uparrow\downarrow> - |\downarrow\uparrow>]$ which cannot be written in terms of the tensor product $|\uparrow> \bigotimes |\downarrow>$ or $|\downarrow> \bigotimes |\uparrow>$. In plain English, the singlet state indicates that one spin is "up" and the other is "down", but it does not tell us which is up and which is down. Each has a 50% probability of being up or down. The tensor product, on the other hand, tells us that the first spin is up and the second spin is down, or the other way around.
Entangled states remain entangled forever. If one of the two electrons went to New York

$$\psi^{11}_{ground} = \frac{2}{\Upsilon_1}|\downarrow\uparrow\uparrow> + \frac{\Lambda_1}{J\Upsilon_1}|\downarrow\uparrow\downarrow> + \frac{2}{\Upsilon_1}|\uparrow\downarrow\downarrow> . \tag{13.6}$$

However, when the two inputs are [1, 1], we want the output to be [0], since this is the situation shown in Fig. 13.1(a) and conforms to the truth table of a NAND gate. Therefore, the *desired* ground state is the unentangled state

$$\psi^{11}_{desired} = |\downarrow\uparrow\downarrow>, \tag{13.7}$$

which corresponds to $A = 1$, $C = 1$ and $B = 0$, as shown in Fig. 13.1(a).

From Equation (13.6) it is obvious that we can make $\psi^{11}_{ground} \approx \psi^{11}_{desired}$ if

$$\left|\frac{\Lambda_1}{2J}\right| = \frac{h + J + \sqrt{(h+J)^2 + 8J^2}}{2J} >> 1, \tag{13.8}$$

i.e., if $h >> J$. In other words, *the ground state will approximate* the state in Fig. 13.1(a) if we make $h >> J$. That means if we apply the inputs [1, 1] to the input dots A and C and let the system relax thermodynamically to the ground state (by emitting phonons, etc.), it will reach the state in Fig. 13.1(a) where the output bit (in dot B) will be [0] and we will have realized the first entry in the truth table of the NAND gate. All that is required is $h >> J$.

Case II: $h_A = h_C =$ **-**$h <$ **0:** This is the case when input bits [0 0] are written in dots A and C by the local magnetic fields. In this case, the eigenenergies and eigenstates are given in Table 13.3.

It is obvious that Table 13.3 can be derived from Table 13.2 by simply replacing h with -h.

The ground state wavefunction is the entangled state

$$\psi^{00}_{ground} = \frac{2}{\Upsilon_1}|\downarrow\uparrow\uparrow> + \frac{\Lambda_1}{J\Upsilon_1}|\uparrow\downarrow\uparrow> + \frac{2}{\Upsilon_1}|\uparrow\uparrow\downarrow>, \tag{13.9}$$

whereas the desired state shown in Fig. 13.1(b) is

$$\psi^{00}_{desired} = |\uparrow\downarrow\uparrow> . \tag{13.10}$$

Once again, we can make $\psi^{00}_{desired} \approx \psi^{00}_{ground}$, if we make $h >> J$. Therefore, if we apply inputs [0 0] to dots A and C and let the system relax to the ground

and the other to Tokyo, and a measurement at Tokyo told us that the spin measured there was down, then it immediately tells us that the spin in New York is up. Information about the New York spin travels to Tokyo instantaneously with infinite speed, apparently violating the tenets of Einstein's relativity. This was the basis of the famous Einstein-Podolsky-Rosen paradox. It has led to deep questions about quantum non-locality and is the basis of Bell's theorem addressing such issues as hidden variables in quantum mechanics. The entangled state is actually used in quantum teleportation. We shall revisit this matter in the next chapter.

TABLE 13.3
Eigenenergies and eigenstates when inputs are [0, 0]

Eigenenergy	Eigenstate
$-J-h+Z-\Delta_1$	$[0,0,0,2/\Upsilon_1,0,\Lambda_1/(J\Upsilon_1),2/\Upsilon_1,0]$
$2J-2h+3Z$	$[0,0,0,0,0,0,0,1]$
$-J+h-Z-\Delta_2$	$[0,2/\Upsilon_3,-\Lambda_3/(J\Upsilon_3),0,2/\Upsilon_3,0,0,0]$
$-Z$	$[0,-1/\sqrt{2},0,0,1/\sqrt{2},0,0,0]$
Z	$[0,0,0,-1/\sqrt{2},0,0,1/\sqrt{2},0]$
$-J-h+Z+\Delta_1$	$[0,0,0,2/\Upsilon_6,0,\Lambda_6/(J\Upsilon_6),2/\Upsilon_6,0]$
$-J+h-Z+\Delta_2$	$[0,2/\Upsilon_7,-\Lambda_7/(J\Upsilon_7),0,2/\Upsilon_7,0,0,0]$
$2J+2h-3Z$	$[1,0,0,0,0,0,0,0]$

state, dot B will have output [1] corresponding to Fig. 13.1(b), and we will have realized the second entry in the truth table of the NAND gate, provided $h >> J$.

Case III: -$h_A = h_C = h > 0$: This is the case when input bits 0 and 1 are written in dots A and C by the local magnetic fields. In this case, the eigenenergies and eigenstates are given in Table 13.4

TABLE 13.4
Eigenenergies and eigenstates when inputs are [0, 1]

Eigenenergy	Eigenstate
$-\Theta_4/2-2J/3-Z+\sqrt{3}i/2\Theta_3$	$[0,\Pi_3^{(1)}/(J^2\Pi_4^{(1)}),2\Pi_1^{(1)}/(J\Pi_4^{(1)}),0,4/\Pi_4^{(1)},0,0,0]$
$-\Theta_4/2-2J/3+Z+\sqrt{3}i/2\Theta_3$	$[0,0,0,\Pi_3^{(2)}/(J^2\Pi_4^{(2)}),0,2\Pi_1^{(2)}/(J\Pi_4^{(2)}),4/\Pi_4^{(2)},0]$
$-\Theta_4/2-2J/3-Z-\sqrt{3}i/2\Theta_3$	$[0,\Pi_3^{(3)}/(J^2\Pi_4^{(3)}),2\Pi_1^{(3)}/(J\Pi_4^{(3)}),0,4/\Pi_4^{(3)},0,0,0]$
$-\Theta_4/2-2J/3+Z-\sqrt{3}i/2\Theta_3$	$[0,0,0,\Pi_3^{(4)}/(J^2\Pi_4^{(4)}),0,2\Pi_1^{(4)}/(J\Pi_4^{(4)}),4/\Pi_4^{(4)},0]$
$2J-3Z$	$[1,0,0,0,0,0,0,0]$
$2J+3Z$	$[0,0,0,0,0,0,0,1]$
$\Theta_4-2J/3-Z$	$[0,\Pi_3^{(7)}/(J^2\Pi_4^{(7)}),2\Pi_1^{(7)}/(J\Pi_4^{(7)}),0,4/\Pi_4^{(7)},0,0,0]$
$\Theta_4-2J/3+Z$	$[0,0,0,\Pi_3^{(8)}/(J^2\Pi_4^{(8)}),0,2\Pi_1^{(8)}/(J\Pi_4^{(8)}),4/\Pi_4^{(8)},0]$

where

$$\Theta_1 = J\left[9(h/J)^2-10+3i\sqrt{3(h/J)^6+12(h/J)^4+69(h/J)^2+27)}\right]^{1/3},$$

$$\Theta_2 = -\frac{4J^2}{3\Theta_1}\left[\left(\frac{h}{J}\right)^2+\frac{7}{3}\right],$$

$$\Theta_3 = \frac{2\Theta_1}{3}+\frac{3\Theta_2}{2} = 2iIm\left(\frac{2\Theta_1}{3}\right),$$

$$\Theta_4 = \frac{2\Theta_1}{3} - \frac{3\Theta_2}{2} = 2Re\left(\frac{2\Theta_1}{3}\right),$$
$$\Pi_1^{(1)} = -\Theta_4/2 - 2J/3 + 2h + (\sqrt{3}i/2)\Theta_3,$$
$$\Pi_2^{(1)} = -\Theta_4/2 - 2J/3 - Z + (\sqrt{3}i/2)\Theta_3,$$
$$\Pi_3^{(1)} = \left[\Pi_2^{(1)}\right]^2 + 2\Pi_2^{(1)}(Z + J + h) + 4Jh + 2JZ - 4J^2 + Z^2 + 2hZ,$$
$$\Pi_4^{(1)} = \left[\frac{[\Pi_3^{(1)}]^2}{J^4} + \frac{4[\Pi_1^{(1)}]^2}{J^2} + 16\right]^{1/2},$$
$$\Pi_1^{(2)} = \Pi_1^{(1)},$$
$$\Pi_2^{(2)} = \Pi_2^{(1)} + 2Z,$$
$$\Pi_3^{(2)} = \left[\Pi_2^{(2)}\right]^2 + 2\Pi_2^{(1)}(-Z + J + h) + 4Jh - 2JZ - 4J^2 + Z^2 - 2hZ,$$
$$\Pi_4^{(2)} = \left[\frac{[\Pi_3^{(2)}]^2}{J^4} + \frac{4[\Pi_1^{(2)}]^2}{J^2} + 16\right]^{1/2},$$
$$\Pi_1^{(3)} = \Pi_1^{(1)} - \sqrt{3}i\Theta_3,$$
$$\Pi_2^{(3)} = \Pi_2^{(1)} - \sqrt{3}i\Theta_3,$$
$$\Pi_3^{(3)} = \left[\Pi_2^{(3)}\right]^2 + 2\Pi_2^{(3)}(Z + J + h) + 4Jh + 2JZ - 4J^2 + Z^2 + 2hZ,$$
$$\Pi_4^{(3)} = \left[\frac{[\Pi_3^{(3)}]^2}{J^4} + \frac{4[\Pi_1^{(3)}]^2}{J^2} + 16\right]^{1/2},$$
$$\Pi_1^{(4)} = \Pi_1^{(2)} - \sqrt{3}i\Theta_3,$$
$$\Pi_2^{(4)} = \Pi_2^{(2)} - \sqrt{3}i\Theta_3$$
$$\Pi_3^{(4)} = \left[\Pi_2^{(4)}\right]^2 + 2\Pi_2^{(4)}(-Z + J + h) + 4Jh - 2JZ - 4J^2 + Z^2 - 2hZ,$$
$$\Pi_4^{(4)} = \left[\frac{[\Pi_3^{(4)}]^2}{J^4} + \frac{4[\Pi_1^{(4)}]^2}{J^2} + 16\right]^{1/2},$$
$$\Pi_1^{(7)} = \Theta_4 - 2J/3 + 2h,$$
$$\Pi_2^{(7)} = \Theta_4 - 2J/3 - Z,$$
$$\Pi_3^{(7)} = \left[\Pi_2^{(7)}\right]^2 + 2\Pi_2^{(7)}(Z + J + h) + 4Jh + 2JZ - 4J^2 + Z^2 + 2hZ,$$
$$\Pi_4^{(7)} = \left[\frac{[\Pi_3^{(7)}]^2}{J^4} + \frac{4[\Pi_1^{(7)}]^2}{J^2} + 16\right]^{1/2},$$
$$\Pi_1^{(8)} = \Pi_1^{(7)},$$
$$\Pi_2^{(8)} = \Pi_2^{(7)} + 2Z,$$

$$\Pi_3^{(8)} = \left[\Pi_2^{(8)}\right]^2 + 2\Pi_2^{(8)}(-Z + J + h) + 4Jh - 2JZ - 4J^2 + Z^2 - 2hZ,$$
$$\Pi_4^{(8)} = \left[\frac{[\Pi_3^{(8)}]^2}{J^4} + \frac{4[\Pi_1^{(8)}]^2}{J^2} + 16\right]^{1/2}, \tag{13.11}$$

where Im stands for the imaginary part, Re stands for the real part, and i is the imaginary square root of -1.

It can be verified that even though the Θ-s are complex, the energy eigenvalues in Table 13.4 are all real, as they must be since they are eigenvalues of a Hermitean matrix.

The ground state wavefunction is the entangled state

$$\psi_{ground}^{01} = \Pi_3^{(1)}/(J^2\Pi_4^{(1)})|\downarrow\downarrow\uparrow> + 2\Pi_1^{(1)}/(J\Pi_4^{(1)})|\downarrow\uparrow\downarrow> + 4/\Pi_4^{(1)}|\uparrow\downarrow\downarrow>, \tag{13.12}$$

whereas the desired state shown in Fig. 13.1(c) is

$$\psi_{desired}^{01} = |\uparrow\downarrow\downarrow> . \tag{13.13}$$

Once again, we can make $\psi_{desired}^{01} \approx \psi_{ground}^{01}$, if we make $h >> J$. Therefore, if we apply inputs [0 1] to dots A and C and let the system relax to the ground state, dot B will have output [1] corresponding to Fig. 13.1(c), and we will have realized the third entry in the truth table of the NAND gate. As before, all we need for this to happen is $h >> J$.

Case IV: -h_A = h_C = -h < 0: This is the case when input bits 1 and 0 are written in dots A and C by the local magnetic fields. In this case, the eigenenergies and eigenstates are given by Table 13.4 with h replaced with -h. The eigenenergies do not change since they depend on h^2 and are therefore insensitive to the sign of h. However, the eigenstates are sensitive to the sign of h, and change.

The ground state wavefunction is

$$\psi_{ground}^{10} = \hat{\Pi}_3^{(1)}/(J^2\hat{\Pi}_4^{(1)})|\downarrow\downarrow\uparrow> + 2\hat{\Pi}_1^{(1)}/(J\hat{\Pi}_4^{(1)})|\downarrow\uparrow\downarrow> + 4/\hat{\Pi}_4^{(1)}|\uparrow\downarrow\downarrow>, \tag{13.14}$$

where we have used the 'hats' for distinction, i.e., to remind the reader that $\hat{\Pi} \neq \Pi$ since the Π-s are sensitive to the sign of h.

The desired state shown in Fig. 13.1(d) is

$$\psi_{desired}^{10} = |\downarrow\downarrow\uparrow> . \tag{13.15}$$

Once again, we can make $\psi_{desired}^{10} \approx \psi_{ground}^{10}$, if we make $h >> J$. Therefore, if we apply inputs [1 0] to dots A and C and let the system relax to the ground state, dot B will have output [1] corresponding to Fig. 13.1(d), and we will have realized the fourth and final entry in the truth table of the NAND gate. All we need is $h >> J$.

13.3.3 Ground state computing with spins

What we have shown in the previous section is that if we provide binary inputs to the two peripheral spins in the 3-spin array using local magnetic fields applied selectively to the input dots, and then let the array relax to the ground state thermodynamically (e.g., by emitting phonons), then we will realize the operation of the NAND gate. This is an important result since the NAND gate is *universal*, meaning that any Boolean logic circuit can be implemented with NAND gates alone. This idea of relaxation to the ground state has some connection with the idea of artificial neural networks, where a system is allowed to relax to ground state to compute.

The ground state computing notion appears in many contexts. One particularly relevant scheme was proposed by Pradeep Bakshi and co-workers in 1991 [17], where they considered an array of 'quantum dashes', each with a single electron. This is shown in Fig. 13.2. Because of Coulomb repulsion, electrons in neighboring dashes are pushed towards opposite corners in antipodal positions. Therefore, the ground state of the array has spontaneous polarization and assumes an anti-ferroelectric configuration, just like our arrays tend to assume an anti-ferromagnetic configuration. Any three dashes will realize a NAND gate [18] in a way very similar to our construct. Two peripheral dashes are input ports and the central dash is the output port. Obviously, when the inputs are [0 0] and [1 1], the output is [1] and [0], respectively. When the inputs are [0 1] or [1 0], there is a tie, which can be resolved in favor of the output bit being [1], by applying a weak global biasing dc electric field, much like our global dc magnetic field.

The ground state computing scheme however has some problems. We now discuss some of them.

13.3.3.1 Error probability

It is the natural tendency of any physical system to relax to the ground state, which is the basis of ground state computing. However, once a system relaxes to ground state, it need not stay there forever. If it gets out of the ground state, and it does [19], it will produce wrong results and cause errors in computation.

Consider a system which is thermodynamically coupled to its environment (phonon bath) which allows it to relax to the ground state, after it is provided with inputs. Once the computation is over and the system has attained equilibrium with the environment, the probability of finding the system in any state (including the ground state) with energy ϵ_n, at a temperature T, is given by the Fermi-Dirac occupation probability $1/exp[(\epsilon_n - E_F)/kT + 1]$ where E_F is the Fermi level. This probability is *not* unity for the ground state. If we approximate the Fermi Dirac statistics with the Boltzmann statistics, then the relative probability of being in two states E_p and E_q is $exp[-(E_p - E_q)/kT]$. If we designate the ground state energy as E_{ground} and an excited state energy as $E_{excited}$, then the ratio of the probabilities of being in the excited state

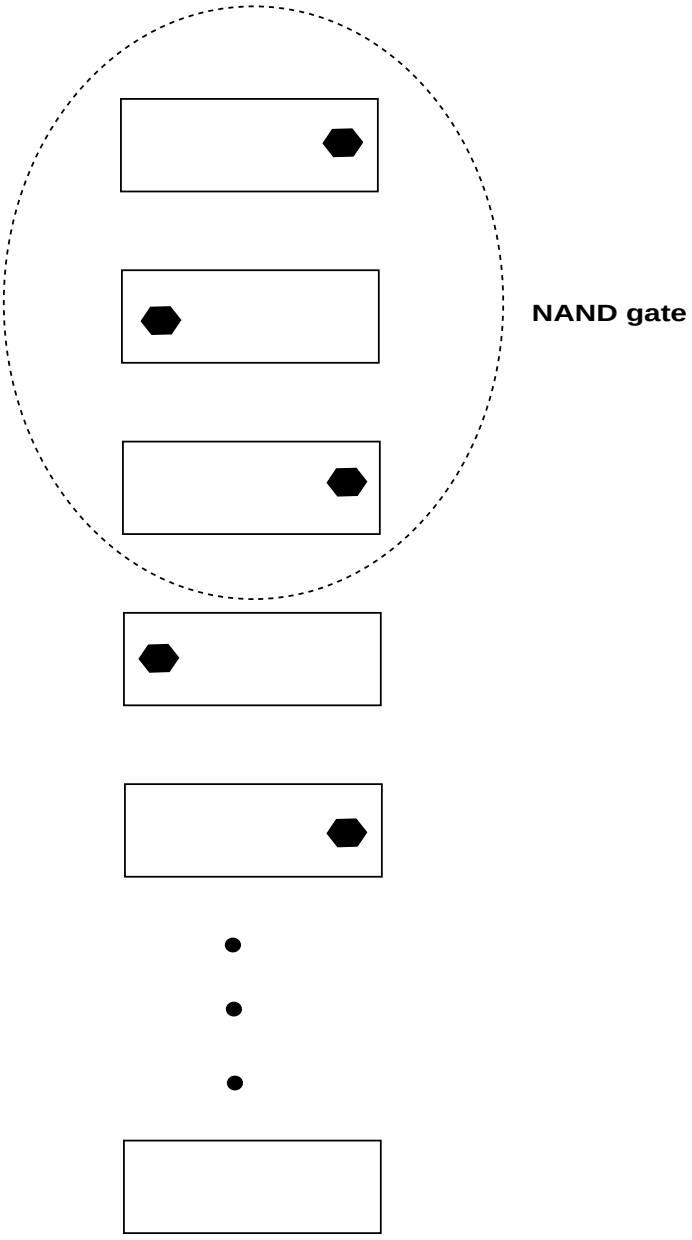


FIGURE 13.2
An array of quantum dashes with nearest neighbor Coulomb coupling. Each dash has a single electron. Coulomb repulsion pushes electrons in neighboring dashes to opposite corners so that the array assumes an anti-ferroelectric order. Three consecutive dashes form a NAND gate. After ref. [17].

and the ground state is $exp[-(E_{excited} - E_{ground})/kT]$. We can call this the *intrinsic error probability* P_{error} associated with being in the excited state, since straying from the ground state into any excited state causes an error in the result. The total error probability is the sum of P_{error} carried out over all excited states.

Case I: inputs are [1 1]: Here $E_1 - E_{ground} \approx 4J - 2Z$ and $E_2 - E_{ground} \approx 2h + 2Z + 2J \approx 2h$ if we assume that $h >> J, Z$. Because of the last inequality, we only have to worry about the first excited state E_1, since the second excited state E_2 is so far above in energy than E_1 that the probability of straying into the second excited state is negligible compared to the probability of straying into the first excited state.

Case II: inputs are [0 0]: Here $E_1 - E_{ground} \approx 4J + 2Z$ and $E_2 - E_{ground} \approx 2h - 2Z + 2J \approx 2h$ if we assume that $h >> J, Z$. Therefore, the same considerations as Case I apply.

Case III and IV: inputs are [1 0] or [0 1]: In these cases also, we need to worry only about the first excited states, since the second excited states are far above in energy than the first excited states. Here, $E_1 - E_{ground} = 2Z$.

The total intrinsic error probability is $P_{error} = p \approx exp[-(E_1 - E_{ground})/kT]$. Therefore, $E_1 - E_{ground} = kTln(1/p)$. Thus, we need two conditions to be fulfilled: (i) $Z = (1/2)kTln(1/p)$, and (ii) $4J - 2Z = 4J - kTln(1/p) = kTln(1/p)$, or $J = (1/2)kTln(1/p)$. Obviously, the required values of $J/(kT)$ and $Z/(kT)$ will be determined by the error probability we can tolerate. Let us say that we can tolerate an error probability of 3% since it is possible to correct for this error rate using elaborate and sophisticated error correction schemes [1] used in quantum computing. Therefore, $J/(kT) = Z/(kT) = (1/2)\ln(1/.03) = 1.75$. Present technology can produce $J \sim 1$ meV with semiconductor quantum dots [20]. Therefore, $T \approx 6.5$ K. Molecular systems can produce $J \sim 6$ meV [21], which raises the temperature of operation to 39 K, that may be attainable with cryocoolers. In order to operate at room temperature, we will need $J = 46$ meV, which is extremely challenging and probably unattainable in the near term.

Handling an error probability of 3% is possible, but requires immense resources. If we cannot afford that, and can only tolerate an error probability of 10^{-9} typical of ultra large scale integrated circuits today, then $T \approx 1.1$ K, as long as $J \sim 1$ meV. Operating at the liquid helium-4 temperature of 4.2 K will require increasing J to about 4 meV, which is possible with molecular systems.

13.3.3.2 The Issue of Unidirectionality

There is however a serious issue in ground state computing which may not be apparent at first. One of the requirements for logic circuits is "unidirectionality", i.e., logic signals must propagate unidirectionally from the input port

to the output port *and not the other way around* [3]. Device engineers call this property "isolation" between the input and output. Traditional transistors have inherent isolation between input and output[§], but in SSL, there is unfortunately no isolation between the input dot and output dot of the logic gate since exchange interaction is bidirectional. Consider just two exchange coupled spins in two neighboring quantum dots. Their ground state will be the singlet state (recall the discussion of parahydrogen in Chapter 9) and therefore this system will act as a natural NOT gate if one spin is the input and the other is the output [the output is always the logic complement of the input]. However, exchange interaction, being bidirectional, cannot distinguish between which spin is the input bit and which is the output. The input will influence the output just as much as the output influences the input. If something spurious or extraneous (such as noise) changes the output, it will change the input! The master-slave relation between input and output is lost. Since the input and output are indistinguishable, it becomes ultimately impossible for logic signal to flow unidirectionally from an input stage to an output stage and not the other way around.

Of course, one can forcibly impose unidirectionality in some (but not all) cases by holding the input spin in a fixed orientation until the desired output spin orientation is produced in the output dot. Once the *final* output has been harvested, it does not matter if things go astray. We can release the input once the final output has been produced. In that case, the input signal itself enforces unidirectionality because it is a symmetry-breaking influence. This approach was actually used to demonstrate a device called a 'magnetic cellular automaton' where the input enforced unidirectionality and produced the correct final output [22]. However, there are problems. First, this approach can only work for a small number of bits before the influence of the input dies out. Second, and more important, the input cannot be changed until the final output has been produced since otherwise the correct output may not be produced at all. That makes such architectures *non-pipelined*[¶] and therefore unacceptably slow. There may also be additional problems associated with random errors when this approach is employed. Because of the lack of unidirectionality, the system can get stuck in metastable states [23] which could prevent relaxation to the ground state. These issues have been discussed in ref. [19].

In our case, we have to hold the inputs in dots A and C at their intended

[§]For example, in SPINFETs, if we change the gate voltage, it will alter the source-to-drain current, but any change in the source-to-drain current does not affect the gate voltage. The interaction is one-way, meaning that there is isolation between the input (gate) terminal and the output (drain) terminal.

[¶]Pipelined architectures are those where a new input can be provided to the computing machinery before the output in response to a previous input has been generated. One does not have to wait for the output to be generated before a new input can be provided. This increases the processing speed vastly. All modern computing architectures allow pipelining.

states until the output in dot B has been produced. In other words, we cannot let go of the inputs in one particular gate until the output in *that* gate has been produced, but we do not have to hold the inputs in place until the outputs in all *succeeding* gates (or, in other words, the final outputs) have been produced. Therefore, pipelining is possible in this approach, if we use some clever scheme to impose unidirectionality. This is discussed next.

13.3.3.3 Unidirectionality in time: clocking

If unidirectionality cannot be imposed in *space* (because exchange interaction is bidirectional), then it must be imposed in *time*. This is accomplished by using clocking to activate successive stages sequentially in time [24]. This strategy is routinely adopted in bucket brigade devices, such as charge-coupled-device (CCD) shift registers [25], where a push clock and a drop clock are used to lower and raise barriers between neighboring charge coupled devices and thus steer a charge packet unidirectionally from one device to the next. In single spin circuits, we can do the same. We will delineate a gate pad between every two neighboring quantum dots and apply a clock signal to this pad. This is shown in Fig. 13.3. During the positive clock cycle, a positive potential will appear over the potential barrier isolating two neighboring quantum dots which will lower the barrier temporarily to make the wavefunctions overlap and exchange couple these two spins. Then during the negative clock cycle, the barrier is raised again to decouple the two spins. In this way, pairs of spin bits can be coupled sequentially in time and logic information transferred unidirectionally from one dot to the next in a bucket brigade fashion. However, there are some subtle issues. Ref. [26] showed that a single phase clock does not work; instead, a 3-phase clock is required to do this job.

13.4 Energy dissipation issues

13.4.1 Energy dissipated in the gate during switching

Consider the ground state energies in Cases I - IV that we considered. These energies are

$$\begin{aligned} E^{I}_{ground} &= -J - h - Z - \Delta_1, \\ E^{II}_{ground} &= -J - h + Z - \Delta_1, \\ E^{III}_{ground} &= E^{IV}_{ground} = -Re\left(\frac{2\Theta_1}{3}\right) - \frac{2J}{3} - Z - \sqrt{3}Im\left(\frac{2\Theta_1}{3}\right), \end{aligned} \tag{13.16}$$

where Re and Im stand for real and imaginary parts.

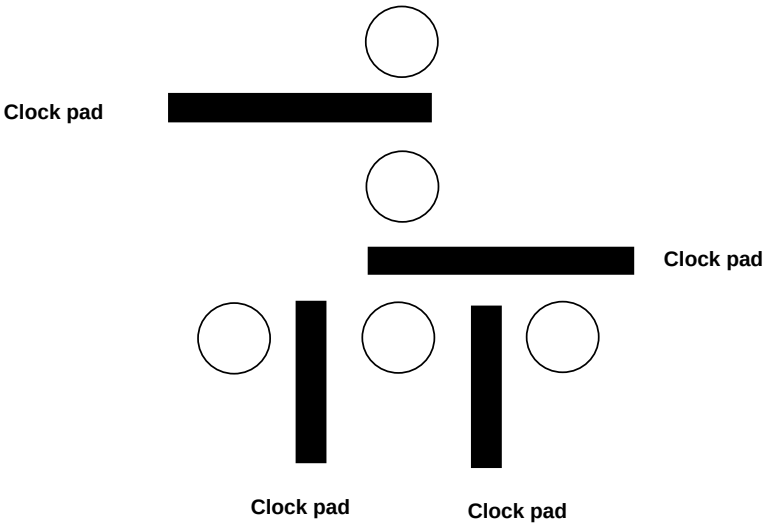


FIGURE 13.3
A clock pad is inserted between neighboring dots. A positive potential appearing at a clock pad will lower the potential barrier between the two flanking dots and cause exchange coupling. A 3-phase clock can propagate signals unidirectionally.

When $h >> J, Z$,

$$E^{I}_{ground} = -2J - 2h - Z,$$
$$E^{II}_{ground} = -2J - 2h + Z,$$
$$E^{III}_{ground} = E^{IV}_{ground} = -2h - 2J/3 - Z. \quad (13.17)$$

Therefore, when we switch between any two of the four states in Fig. 13.1, the maximum energy that can be dissipated is $2Z$. We saw that this energy is equal to $kTln(1/p)$. Therefore, the maximum energy dissipated in switching the NAND gate is $kTln(1/p)$. With $p = 10^{-9}$, it is $\sim$ 21kT, whereas modern electronic transistors dissipate 40,000 - 50,000kT during switching. At a temperature of 1 K, the energy dissipated is 3×10^{-22} Joules. With a 50 GHz clock, the corresponding dissipated power is about 15 picoWatts/gate. This will allow us to continue to shrink electronic devices well into the next few decades, i.e., continue the fabled Moore's law[‖] without having to worry about excessive power dissipation and thermal management on the chip.

The dissipated energy of $kTln(1/p)$ is actually the minimum energy that needs to be dissipated to switch a NAND gate if we do not wish to exceed an error probability of p. This is the famous *Landauer-Shannon result* [27] which stipulates that in a *logically irreversible* operation like the NAND operation, the energy dissipated is $kTln(1/p)$, where p is the error probability. If $p > 0.5$, then the distinguishability between bits 0 and 1 is lost. Therefore, the absolute minimum energy that must be dissipated to switch any irreversible logic gate is $kTln2$.

Landauer also showed that it is possible to switch *logically reversible* gates without dissipating any energy at all, but at the expense of losing error correction capability [27]. Logically reversible gates are those where the inputs can be inferred from a knowledge of the outputs, meaning that the logic function implemented by the gate has a unique inverse. An inverter is logically reversible, since the input is always the logic complement of the output. Therefore, knowledge of the output is sufficient to tell us unambiguously what the input was. However, a NAND gate is logically irreversible. When the output is [0], we can tell precisely what the inputs were since they must have been [1 1]. However, when the output is [1], we cannot tell with certainty what the inputs were. They could have been [0 0], or [0 1], or [1 0]. Therefore, a NAND gate is logically irreversible. There is no way to switch such a gate without dissipating energy.

[‖]Moore's law is an empirical law postulated by Gordon Moore, one of the founders of Intel Corporation, which states that the device density on a chip will roughly double every 18 months. In order for this to remain feasible, the energy dissipation per device should also reduce by approximately 50% every 18 months. Otherwise, the power dissipation (per unit area) on the chip will continue to increase and stretch the limits of heat sinking technology.

Exercise: A simple proof of the Landauer-Shannon result

The energy dissipated in changing the state of a system, which is in equilibrium with the environment (phonon bath), is

$$\Delta E = kT\Delta S + S\Delta(kT), \tag{13.18}$$

where S is the entropy.

According to the second law of thermodynamics, $\Delta(kT) \geq 0$, so that

$$\Delta E_{min} = kT\Delta S. \tag{13.19}$$

Shannon's definition of entropy is

$$S = \sum_{i}^{N} p_i ln(p_i), \tag{13.20}$$

where the system in question can exist in N states and p_i is the probability of being in the i-th state.

A binary system has only two states ($i = 0, 1$) and $N = 2$. Therefore,

$$S = p_{[0]} ln\left(p_{[0]}\right) + p_{[1]} ln\left(p_{[1]}\right). \tag{13.21}$$

After switching the system, let us say that we find it is in state [1] so that $p_{[1]}^{final} = 1$ and $p_{[0]}^{final} = 0$. Therefore,

$$S_{final} = 0ln(0) + 1ln(1) = 0. \tag{13.22}$$

If the gate is *logically irreversible* and we could not infer the input state from knowledge of the output, then $p_{[0]}^{initial} = p_{[1]}^{initial} = 0.5$. In that case,

$$S_{final} = 0.5ln(0.5) + 0.5ln(0.5) = -ln(2). \tag{13.23}$$

Therefore, $\Delta S = S_{final} - S_{initial} = ln(2)$. Substituting this in Equation (13.19), we get the Landauer-Shannon result:

$$E_{min} = kTln2. \tag{13.24}$$

If the gate were *logically reversible*, i.e., we could infer the input from the output, then the situation will be different. Consider an inverter. If the output were bit 1, then input must have been bit 0. In that case, $S_{initial} = 1ln(1) + 0ln(0) = 0$, and $S_{final} = 0ln(0) + 0ln(0) = 0$. Therefore, $\Delta S = S_{final} - S_{initial} = 0$, and $\Delta E_{min} = 0$, meaning dissipationless switching is possible.

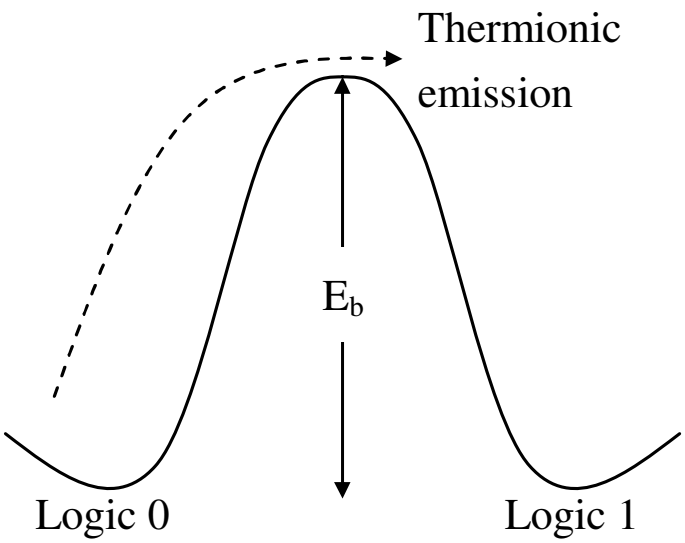


FIGURE 13.4
The potential energy landscape of a bistable device. An energy barrier E_b separates the two potential minima which 'store' the two states. The barrier is high and/or wide enough to suppress tunneling. Thermionic emission over the barrier can cause spontaneous transitions between the two states resulting in bit errors.

Physical interpretation of the Landauer-Shannon result

A bistable element used to store and process binary logic bits must have two stable (or metastable) states. Viewed from the perspective of an energy landscape, these two states will form two potential minima as shown in Fig. 13.4. In order to prevent spontaneous transitions between these two states, which will cause bit errors, there must be an energy barrier separating these two minima. When the logic element is to be switched, the potential barrier must be lowered by some external agency so that one state becomes accessible from the other via elastic transitions.

Based on the second law of thermodynamics, the amount of energy dissipated in lowering the barrier is at least equal to the height of the barrier. Therefore, the minimum energy dissipated in switching is the minimum barrier height that we have to maintain in order to suppress spontaneous transitions between the two potential wells. These transitions are caused by thermionic emission over the barrier. The probability of thermionic emission over a bar-

rier of height E_b is $exp[-E_b/kT]$ assuming that the carriers are distributed in a Boltzmann distribution characterized by a carrier temperature T. If this probability approaches 0.5, then it is as much likely that the logic element is in one state as the other, meaning that there is no preference to be in either state. The distinguishability of the two states will then be completely lost. Therefore, we need

$$e^{-\frac{E_b}{kT}} \leq 0.5 \tag{13.25}$$

or $E_b > kTln2$. This is the physical basis of the Landauer-Shannon limit on the minimum energy dissipated during switching.

It is obvious that there are many assumptions implicit in Equation (13.25). For example, we assumed that the carriers are in a Boltzmann distribution. They need not be. If they are driven far out of equilibrium, then they could be in a non-thermal distribution and Equation (13.25) will not apply. This situation has been examined in ref. [28] which reached the conclusion that it may be possible to perform irreversible logic operations while dissipating less than $kTln2$ amount of energy per bit flip, using non-equilibrium systems.

13.4.2 Energy dissipated in the clocking circuit

The clocking circuit that one needs to impose unidirectionality in SSL introduces additional dissipation. The energy dissipated in the clock pad is $(1/2)CV^2$ if the clock pulse is applied abruptly. It can be much less if applied in small steps (adiabatically) [29], but we will not consider that scenario here since it requires extreme precision. Appropriate pulse shaping can also reduce this energy dissipation, but usually at the expense of slower speed. Since we want to be conservative, we will assume the maximum dissipation of $(1/2)CV^2$. Here, C is the capacitance of the gate pad and V is the potential applied to the pad to lower the potential barrier between neighboring dots to exchange couple neighboring spins. The value of V may be as small as 10 mV.

13.5 Comparison between hybrid and monolithic spintronics

For the sake of comparison, we will assume that the capacitance of a gate pad in SSL is about the same as the gate capacitance of a SPINFET, which was estimated in the previous chapter to be 50 aF. We will carry out a comparison at 4.2 K, since that is the order of temperature required to achieve high spin injection efficiency in a SPINFET.

The energy dissipated in switching a single gate in SSL is $kTln(1/p)$. With an error probability of $p = 10^{-9}$ (achieved in modern integrated circuits), this energy is 7.5 meV at 4.2 K. We require that J, Z is one half of this, which is 3.75 meV. It is not impossible to achieve this magnitude of J [20]. In order to make $Z = 3.75$ meV, we need a magnetic field of only 0.07 Tesla if we use materials with giant g-factors (e.g., $InSb_xN_{1-x}$, which is reported to have a g-factor of 900 [30]). Of course, these materials may not be optimum for long spin relaxation times. Even if we can find materials with g-factors of ~ 10 that have long spin relaxation times, the required magnetic field is about 6 Tesla, which is easily achievable.

The energy dissipated in the clock is 2.5×10^{-21} Joules, while the energy dissipated in the gate is 7.5 meV $= 1.2 \times 10^{-21}$ Joules. Therefore, the total energy dissipated in an SSL gate during switching is $\sim 4 \times 10^{-21}$ Joules which is much better than the $\sim 2 \times 10^{-16}$ Joules that we computed for the SPINFET in Section 12.4. While these numbers depend somewhat on device parameters that were assumed, it is safe to assert that SSL – an example of monolithic spintronics – will be able to reduce power dissipation far more than SPINFETs which are examples of hybrid spintronics.

13.6 Concluding remarks

Recent advances in controlling and manipulating single electrons in quantum dots [31, 32, 33, 34, 35, 36, 37, 38] have now brought us to the point where a paradigm like SSL may be around the corner. Practical implementations may still be several years away, but there is hope that monolithic spintronics may some day be able to compete with the behemoth of traditional silicon based electronics in mainstream applications.

13.7 Problems

- **Problem 13.1**

  There are two sources of bit flip errors: (i) spins flipping spontaneously within a clock period due to spin relaxation mechanisms discussed in Chapter 8. This error probability is $p_1 = 1 - e^{-T/T_1}$, where T is the refresh clock period and T_1 is the spin relaxation time, and (ii) the system getting out of the ground state to the first excited state (we ignore the second and higher excited states since they are too far away in energy). That probability is $p_2 = exp[-(E_1 - E_{ground})/kT]$. At what

clock frequency will the two probabilities be equal if spin relaxation time is 10 msec, the global magnetic flux density is 4 Tesla and the quantum dot material has a g-factor of 20?

- **Problem 13.2**

  Let us say that the up-spin state is the +z-polarized state $[1\ 0]^{\dagger}$ and the down-spin state is the -z-polarized state $[0\ 1]^{\dagger}$. Show that

$$\begin{aligned}
\sigma_z|\uparrow> &= |\uparrow>, \\
\sigma_z|\downarrow> &= -|\downarrow>, \\
J\sigma_{zA}\sigma_{zB}|\uparrow_A\downarrow_B> &= -J|\uparrow_A\downarrow_B>, \\
J\sigma_{zA}\sigma_{zB}|\uparrow_A\uparrow_B> &= J|\uparrow_A\uparrow_B>, \\
\sigma_x|\uparrow> &= |\downarrow>, \\
\sigma_x|\downarrow> &= |\uparrow>, \\
\sigma_y|\uparrow> &= i|\downarrow>, \\
\sigma_y|\downarrow> &= -i|\uparrow>, \\
J\sigma_{xA}\sigma_{xB}|\uparrow_A\downarrow_B> &= J|\downarrow_A\uparrow_B>, \\
J\sigma_{yA}\sigma_{yB}|\uparrow_A\downarrow_B> &= J|\downarrow_A\uparrow_B>, \\
J\sigma_{xA}\sigma_{xB}|\uparrow_A\uparrow_B> &= J|\downarrow_A\downarrow_B>, \\
J\sigma_{yA}\sigma_{yB}|\uparrow_A\uparrow_B> &= -J|\downarrow_A\downarrow_B> . \qquad (13.26)
\end{aligned}$$

- **Problem 13.3**

  Derive the matrix in Equation (13.3) starting with the Heisenberg Hamiltonian.

- **Problem 13.4**

  If we changed the designation so that the down-spin state became logic 0 and the up-spin state logic 1, will we have realized the AND gate? Explain.

- **Problem 13.5**

  We showed that the NAND gate functionality is realized if $h >> J$. The question is how much larger should h be compared to J. Show that it is sufficient that $h \geq 10J$.

  Hint

  Consider the case when the inputs are [0 0] or [1 1]. The requirement that the ground state wavefunction becomes the desired wavefunction is that $|\Lambda_1/(J\Upsilon_1)|^2 \rightarrow 1$ and $|2/\Upsilon_1|^2 \rightarrow 0$. Plot these two quantities as a function of h/J and show that $|\Lambda_1/(J\Upsilon_1)|^2$ begins to saturate towards 1 and $|2/\Upsilon_1|^2$ begins to saturate towards 0 when h/J exceeds 10.

- **Problem 13.6**

  Show that when the inputs are [1 0] or [0 1], it is still sufficient that $h \geq 10J$ in order to yield NAND gate functionality.

- **Problem 13.7**

  We showed that $Z = g\mu_B B_{global} = (1/2)kTln(1/p)$ where B_{global} is the flux density of the global magnetic field. Let us say that we are operating at $T = 1$ Kelvin and $p = 10^{-9}$. What is the value of B_{global} required if the g-factor of the quantum dots hosting the spins is 50?

  What would be the value of B_{global} if we could operate at room temperature with an error probability of 3%?

  Are these reasonable values?

- **Problem 13.8**

  An SSL circuit is operating at $T = 1$K and with bit error probability of $p = 10^{-9}$. The quantum dot material has a g-factor of -50. What is the minimum magnetic field B_{input} required to write input data, i.e., orient the input spins in the desired direction?

  Solution

  We know $J = Z = (1/2)kTln(1/p)$. Therefore, $J = Z = 0.9$ meV, which is quite achievable with present day quantum dot technology. We also saw in Problems 13.5 and 13.6 that the minimum value of h that we require for NAND gate functionality is $10J$. Therefore, $h = |g|\mu_B B_{input} \geq 9$ meV, which yields that the minimum value of $B_{input} = 3.1$ Tesla.

13.8 References

[1] E. Knill, "Quantum computing with realistically noisy devices", Nature (London), **434**, 39 (2005).

[2] S. Salahuddin and S. Datta, "Interacting systems for self correcting low power switching", Appl. Phys. Lett., **90**, 093503 (2007).

[3] D. A. Hodges and H. G. Jackson, *Analysis and Design of Digital Integrated Circuits*, 2nd edition, McGraw Hill, New York, 1988, Chapter 1, p. 2.

[4] M. Ikezawa, B. Pal, Y. Masumoto, I. V. Ignatiev, S. Yu. Verbin and I. Ya. Gerlovin, "Sub millisecond electron spin relaxation in InP quantum dots", Phys. Rev. B, **72**, 153302 (2005).

[5] S. Amasha, K. MacLean, Iuliana Radu, D. M. Zümbuhl, M. A. Kastner, M. P. Hanson and A. C. Gossard, "Electrical control of spin relaxation in a quantum dot", arXiv:0707.1656.

[6] S. Pramanik, C-G Stefanita, S. Patibandla, S. Bandyopadhyay, K. Garre, N. Harth and M. Cahay, "Observation of extremely long spin relaxation time in an organic nanowire spin valve", Nature Nanotech, **2**, 216 (2007).

[7] S. Bandyopadhyay, "Power dissipation in spintronic devices: A general perspective", J. Nanosci. Nanotech., **7**, 168 (2007).

[8] D. Rugar, R. Budakian, H, J. Mamin and B. H. Chui, "Single spin detection by magnetic resonance force microscopy", Nature (London), **430**, 329 (2004).

[9] J. M. Elzerman, R. Hanson, L. H. Willems van Beveren, B. Witkamp, L. M. K. Vandersypen and L. P. Kouwenhoven, "Single shot read out of an electron spin in a quantum dot", Nature (London), **430**, 431 (2004).

[10] M. Xiao, I. Martin, E. Yablonovitch and H. W. Jiang, "Electrical detection of the spin resonance of a single electron in a silicon field effect transistor", Nature (London), **430**, 435 (2004).

[11] S. Bandyopadhyay, B. Das and A. E. Miller, "Supercomputing with spin polarized single electrons in a quantum coupled architecture", Nanotechnology, **5**, 113 (1994).

[12] S. K. Sarkar, T. Bose and S. Bandyopadhyay, "Single spin logic circuits", Phys. Low Dim. Struct., **2**, 69 (2006).

[13] N. W. Ashcroft and N. D. Mermin, *Solid State Physics*, Saunders College, Philadelphia, 1976.

[14] S. N. Molotkov and S. S. Nazin, "Single electron spin with quantum dot logical gates with ferromagnetic chains", Phys. Low Dim. Struct., **10**, 85 (1997).

[15] S. N. Molotkov and S. S. Nazin, "Single electron spin logical gates", JETP Lett., **62**, 256 (1995); S. N. Molotkov and S. S. Nazin, "Single electron computing: Quantum dot logic gates", Zh. Eksp. Teor. Fiz. **110**, 1439 (1996); [Sov. Phys.: JETP, **83**, 794 (1996)].

[16] H. Agarwal, S. Pramanik and S. Bandyopadhyay, "Single spin universal Boolean logic gate", New J. Phys., **10**, 015001 (2008).

[17] P. Bakshi, D. Broido and K. Kempa, "Spontaneous polarization of electrons in quantum dashes", J. Appl. Phys., **70**, 5150 (1991); K. Kempa, D. A. Broido and P. Bakshi, "Spontaneous polarization in quantum dot systems", Phys. Rev. B, **43**, 9343 (1991).

[18] P. Bakshi, private communication, 1992.

[19] M. Anantram and V. P. Roychowdhury, "Metastable states and information propagation in a one-dimensional array of locally coupled bistable cells", J. Appl. Phys., **85**, 1622 (1999).

[20] D. V. Melnikov and J-P Leburton, "Single particle state mixing in two electron double quantum dots", Phys. Rev. B, **73**, 155301 (2006); J-P Leburton, private communication.

[21] C. F. Hirjibehedin, C. P. Lutz and A. G. Heinrich, "Spin coupling in engineered atomic structures", Science, **312**, 1021 (2006).

[22] R. P. Cowburn and M. E. Welland, "Room temperature magnetic quantum cellular automata", Science, **287**, 1466 (2000).

[23] R. Landauer, "Is quantum mechanics useful?", Philos. Trans. Royal Soc. London, Ser. A, **353**, 367 (1995).

[24] S. Bandyopadhyay and V. P. Roychowdhury, "Computational paradigms in nanoelectronics: Quantum coupled single electron logic and neuromorphic networks", Jpn. J. Appl. Phys., **35**, Part 1, 3350 (1996).

[25] D. K. Schroeder, *Advanced MOS Devices*, Modular Series in Solid State Devices, Eds. G. W. Neudeck and R. F. Pierret, (Addison-Wesley, Reading, MA, 1987).

[26] S. Bandyopadhyay, "Computing with spins: From classical to quantum computing", Superlat. Microstruct., **37**, 77 (2005).

[27] R. W. Keyes and R. Landauer, "Minimal energy dissipation in logic", IBM J. Res. Develop., **14**, 152 (1970); R. Landauer, "Uncertainty principle and minimal energy dissipation in the computer", Int. J. Theor. Phys., **21**, 283 (1982).

[28] R. K. Cavin, V. V. Zhirnov, G. I. Bourianoff, J. A. Hutchby, D. J. C. Herr, H. H. Hosack, W. H. Joyner and T. A. Woolridge, "A long term view of research targets in nanoelectronics", J. Nanoparticle Res., **7**, 573 (2005); V. V. Zhirnov, R. K. Cavin, J. A. Hutchby and G. I. Bourianoff, "Limits to binary logic switch scaling - A Gedanken model", Proc. IEEE, **91**, 1934 (2003).

[29] R. K. Cavin, V. V. Zhirnov, J. A. Hutchby and G. I. Bourianoff, "Energy barriers, demons and minimum energy operations of electronic devices", Fluctuation and Noise Letters, **5**, C29 (2005).

[30] X. W. Zhang, W. J. Fan, S. S. Li and J. B. Xia, "Giant and zero g-factors of dilute nitride semiconductor nanowires", Appl. Phys. Lett., **90**, 193111 (2007).

[31] M. Ciorga, A. S. Sachrajda, P. Hawrylak, C. Gould, P. Zawadzki, S. Jullian, Y. Feng and Z. Wasilewski, "Addition spectrum of a lateral

quantum dot from Coulomb and spin blockade spectroscopy", Phys. Rev. B, **61**, R16315 (2000).

[32] M. Piero-Ladriere, M. Ciorga, J. Lapointe, P. Zawadzki, M,. Korukusinski, P. Hawrylak and A. S. Sachrajda, "Spin blockade spectroscopy of a two level artificial molecule", Phys. Rev. Lett., **91**, 026803 (2003).

[33] C. Livermore, C. H. Crouch, R. M. Westervelt, K. L. Campman and A. C. Gossard, "The Coulomb blockade in coupled quantum dots", Science, **274**, 1332 (1996).

[34] T. H. Oosterkamp, T. Fujisawa, W. G. van der Wiel, K. Ishibashi, R. V. Hijman, S. Tarucha and L. P. Kouwenhoven, "Microwave spectroscopy of a quantum dot molecule", Nature (London), **395**, 873 (1998).

[35] A. W. Holleitner, R. H. Blick, A. K. Huttel, K. Eberl and J. P. Kotthaus, "Probing and controlling the bonds of an artificial molecule", Science, **297**, 70 (2001).

[36] N. J. Craig, J. M. Taylor, E. A. Lester, C. M. Marcus, M. P. Hanson and A. C. Gossard, "Tunable non-local spin control in a coupled quantum dot system", Science, **304**, 565 (2004).

[37] R. Hanson, B. Witkamp, L. M. K. Vandersypen, L. H. W. van Beveren, J. M. Elzerman and L. P. Kouwenhoven, "Zeeman energy and spin relaxation in a one electron quantum dot", Phys. Rev. Lett., **91**, 196802 (2003).

[38] J. R. Petta, A. C. Johnson, J. M. Taylor, E. A. Laird, A. Yacoby, M. D. Lukin, C. M. Marcus, M. P. Hanson and A. C. Gossard, "Coherent manipulation of single electron spins in coupled quantum dots", Science, **309**, 2180 (2005).

14

Quantum Computing with Spins

In the last chapter, we saw that it is theoretically possible to switch a *logically reversible* gate without dissipating any energy at all (recall the proof of the Landauer-Shannon result). If there is no dissipation, then no information is ever being discarded; therefore, if we know the state of the system completely at any instant of time, we should be able to infer its state at any previous instant. This is consistent with the equations of quantum mechanics (e.g., the Schrödinger equation, or the Pauli Equation) which are reversible in time. This property of quantum mechanical systems, known as 'time reversal symmetry', guarantees that if we know the state of a system at a time t_0, we can tell what the state was at any time $t < t_0$. Because of this fundamental connection between physical reversibility, logical reversibility, and the equations of quantum mechanics, we would expect logically reversible gates to behave as non-classical (quantum mechanical) computing machinery*. This will bring us to the topic of quantum computing and quantum logic gates, but before we get there, let us explore reversible logic in a little more detail.

14.1 The quantum inverter

The inverter, or the NOT gate, is a logically reversible gate since whenever we know the output of the gate, we can tell unambiguously what the input was. After all, the input is always the logic complement of the output. Since this gate is logically reversible, it should be possible to operate it without dissipating energy. Taking it one step further, we should be able to switch this gate in accordance with the laws of quantum mechanics. This brings us to the topic of a *quantum inverter*.

The quantum inverter is a NOT gate where the output is always the logic complement of the input. Therefore, if we change the input, the output should change accordingly to become the logic complement of the input. In a quantum inverter, this change should take place without dissipating energy and

*Not all reversible gates are quantum mechanical. Classical gates can also be reversible as we will see in the example of the Toffoli-Fredkin gate.

in accordance with the laws of quantum mechanics, i.e., following a unitary evolution.

A quantum inverter can be implemented with two exchange coupled spins. One spin will represent the input bit and the other will be the output bit. Input-output relations will be determined by exchange coupling between the spins, in much the same way as the NAND gate of the previous chapter, but with one major difference. The dynamics of the NAND gate was dissipative. In fact, dissipation is required to make the system relax to the ground state and produce the correct result. In the case of the inverter, there will be no dissipation. Therefore, we will have to rely entirely on the machinery of quantum mechanics to produce the correct result every time. We will show that quantum mechanics is up to the task, but we shall require extreme precision. We must allow the system to evolve according to quantum mechanical laws up to a certain time and then abruptly halt the system, at *precisely the right juncture*, if we want the correct result. If we overshoot or undershoot in time, we will not get the correct answer. This is the Achilles' heel of quantum gates. They are not forgiving of imperfection because there is no error correction capability. Error correction requires dissipation which is forbidden for quantum gates[†]. Therefore, if we have errors, they will build up with time and ultimately render the gate useless. That is why precision is vital for quantum gates.

The NAND gate did not require this kind of precision. All we had to do was to let the system relax to the ground state by dissipating energy. We did not have to bother how long it took or what route the system traced in order to reach the ground state. Those details were irrelevant. There was no issue of overshooting or undershooting. Dissipation forgave all our transgressions and made the gate work in spite of imprecisions and inaccuracies. Unfortunately, we do not have that luxury when we deal with dissipationless systems. Therefore, reversibility comes with a price, namely the need for extreme precision.

Consider the quantum inverter shown in Fig. 14.1. The spin polarization in one dot (say, the left dot) is the input bit and the spin polarization in the other dot is the output bit. We will assume that there is only one size quantized level in each dot. A weak magnetic field B_z is applied globally to define the spin quantization axes, meaning that the spin in any dot is always either parallel or anti-parallel to this field. We will assume that the field is in the z-direction.

The Hubbard Hamiltonian for this system is [1]

[†]Errors can be corrected with 'software', such as error correcting codes, but not by 'hardware'. There are quantum control techniques to minimize errors, but they are not exactly hardware correction approaches.

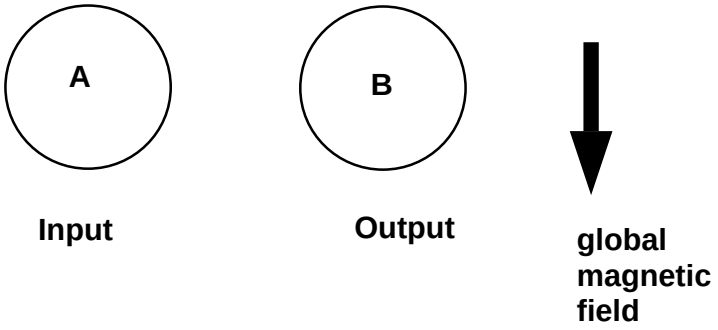


FIGURE 14.1
A quantum inverter realized with two exchange coupled spins in quantum dots. A is the input and B the output of the inverter.

$$H_{Hubbard} = \sum_{i\sigma} \epsilon_0 n_{i\sigma} + \sum_{input\ dot,\ \sigma} g\mu_B B_i sign(\sigma) + \sum_{<ij>} \left[c^{\dagger}_{i\sigma} c_{j\sigma} + h.c. \right] + \sum_{i} U_i n_{i\uparrow} n_{i\downarrow} + \sum_{<ij>\alpha\beta} J_{ij} c^{\dagger}_{i\alpha} c_{i\beta} c^{\dagger}_{j\beta} c_{j\alpha} + B_z \sum_{i\sigma} g\mu_B n_{i\sigma} sign(\sigma) \ , \tag{14.1}$$

where *h.c.* stands for 'Hermitean conjugate'. Here the first term represents the site energy (kinetic + potential) in the i-th dot, the second term is the Zeeman splitting energy associated with applying a local magnetic field to the i-th dot to selectively align the spin in that dot along the direction of the applied local field (providing input data), the third term is the tunneling between nearest neighbor dots (the angular braket denotes 'nearest neighbors'), the fourth term is the charging energy of the dot, which is the energy cost associated with putting a second electron in the same dot (Pauli exclusion principle mandates that the second electron must have a spin anti-parallel to that of the first, since we assume that each dot has only one bound state), the fifth term is the exchange interaction between nearest neighbor dots and the last term is the Zeeman energy due to the global magnetic field, applied over the entire 2-spin array, in the z-direction, to define the spin quantization axes.

Once again, we will simplify this to the Heisenberg model which yields

$$\mathcal{H} = J \sum_{<ij>} \sigma_{zi}\sigma_{zj} + J \sum_{<ij>} [\sigma_{xi}\sigma_{xj} + \sigma_{yi}\sigma_{yj}] + \sum_{input\ dot} \sigma_{zi} h_{zi}^{input} + \sum_{i} \sigma_{zi} h_{zi}^{global}. \tag{14.2}$$

The quantity J is the exchange coupling energy, h_{zi}^{input} is the Zeeman splitting caused by the local magnetic field applied to the left dot (input) to orient the spin of its electron, and h_{zi}^{global} is the Zeeman splitting associated with the global magnetic field.

We will define "up-spin" as the polarization parallel to the global magnetic field and "down-spin" as anti-parallel. In the basis of two electron states, the Hamiltonian in Equation (14.2) can be written as

$$\begin{array}{c} \begin{array}{cccc} \quad|\downarrow\downarrow> & \quad|\downarrow\uparrow> & \quad|\uparrow\downarrow> & \quad|\uparrow\uparrow> \end{array} \\ \begin{pmatrix} h_A + J + 2Z & 0 & 0 & 0 \\ 0 & h_A - J & 2J & 0 \\ 0 & 2J & -h_A - J & 0 \\ 0 & 0 & 0 & -h_A + J - 2Z \end{pmatrix} \begin{array}{l} |\downarrow\downarrow> \\ |\downarrow\uparrow> \\ |\uparrow\downarrow> \\ |\uparrow\uparrow> \end{array} \end{array} \tag{14.3}$$

where $2h_A$ is the Zeeman splitting caused by the externally applied local magnetic field in input dot A and $2Z$ is the Zeeman splitting caused in any dot by the global magnetic field. If the local magnetic field providing input to dot A aligns the spin in the direction of the global magnetic field (up-spin direction), then the corresponding h is positive. Otherwise, it is negative. The quantity J is always positive to ensure that the singlet state has lower energy than the triplet state.

The eigenenergies E_n and eigenstates ϕ_n $(= c_1|\downarrow\downarrow> + c_2|\downarrow\uparrow> + c_3|\uparrow\downarrow> + c_4|\uparrow\uparrow>)$ of the above Hamiltonian are found by diagonalizing the above matrix:

TABLE 14.1
Eigenenergies and eigenstates of the quantum inverter

Eigenenergies	Eigenstates
$-J - \sqrt{h_A^2 + 4J^2}$	$\sqrt{\frac{1}{2}\left(1 - \frac{h_A}{\sqrt{h_A^2+4J^2}}\right)}\|\uparrow\downarrow> - \sqrt{\frac{1}{2}\left(1 + \frac{h_A}{\sqrt{h_A^2+4J^2}}\right)}\|\downarrow\uparrow>$
$-h_A + J$	$\|\uparrow\uparrow>$
$-J + \sqrt{h_A^2 + 4J^2}$	$\sqrt{\frac{1}{2}\left(1 + \frac{h_A}{\sqrt{h_A^2+4J^2}}\right)}\|\uparrow\downarrow> + \sqrt{\frac{1}{2}\left(1 - \frac{h_A}{\sqrt{h_A^2+4J^2}}\right)}\|\downarrow\uparrow>$
$h_A + J + 2Z$	$\|\downarrow\downarrow>$

where, once again, we have arranged the eigenstates in ascending order

(ground state first and highest excited state last).

In the absence of any input ($h_A = 0$), the ground state energy is -3J and the ground state wave function is $\frac{1}{\sqrt{2}}(|\uparrow\downarrow> - |\downarrow\uparrow>)$ which is an entangled state. In this state, neither the input dot nor the output dot has a definite spin polarization.

If the system is initially without any input and is in the ground state at time $t = 0$, then obviously $c_1(0) = c_4(0) = 0$, $c_2(0) = -1/\sqrt{2}$ and $c_3(0) = 1/\sqrt{2}$. Thereafter, if a local magnetic field is applied to the input dot A at time t=0 to align its spin polarization to the "down" state, then in the absence of any coupling to the environment, the wavefunction of the system evolves unitarily according to

$$\psi(t) = exp[-i\mathcal{H}t/\hbar]\psi(0), \tag{14.4}$$

where $\psi(0) = \frac{1}{\sqrt{2}}(|\uparrow\downarrow> - |\downarrow\uparrow>)$.

The wavefunction at time t is given by

$$\psi(t) = c_2(t)|\downarrow\uparrow> + c_3(t)|\uparrow\downarrow>, \tag{14.5}$$

where

$$\begin{aligned} c_2(t) &= -\frac{e^{iJt/\hbar}}{\sqrt{2}}\left[cos(\omega t) + i\left(\frac{h_A}{\hbar\omega} + \sqrt{1 - \frac{h_A^2}{\hbar^2\omega^2}}\right) sin(\omega t)\right], \\ c_3(t) &= \frac{e^{iJt/\hbar}}{\sqrt{2}}\left[cos(\omega t) - i\left(\frac{h_A}{\hbar\omega} - \sqrt{1 - \frac{h_A^2}{\hbar^2\omega^2}}\right) sin(\omega t)\right], \end{aligned} \tag{14.6}$$

and $\hbar\omega = \sqrt{h_A^2 + 4J^2}$.

If the system were to act as an inverter, it must, at some point in time, reach the unentangled state $|\downarrow\uparrow>$. In that state, the left dot (dot A) will be in the down-spin state because of the input provided to it, and the right dot will be in the up-spin state, thereby satisfying the requirement of the inverter, namely that the output is the logic complement of the input. This desired state corresponds to $|c_2| = 1$ and $c_3 = 0$. It is not clear at the outset whether the system can *ever* reach the desired state by evolving unitarily according to Equation (14.4). It turns out (see Problem 14.1) that it can after a time delay

$$\tau_d = \frac{h}{4\sqrt{h_A^2 + 4J^2}} \tag{14.7}$$

provided $h_A = 2J$.

This desired state however is not an eigenstate of the system (see Table 14.1). Consequently, the system will continue to evolve to a different state unless a "read" operation collapses the wave function as soon as the desired state is reached. There is no margin of error allowed. We must "halt" the system as soon as the duration τ_d elapses. If we undershoot or overshoot, we will incur an error that will build up with time. These errors can only be

corrected with 'software', since there are no 'hardware' antidotes. Amazing advances however have been made in 'software', and as mentioned in the previous Chapter, error correction codes exist that can correct errors occurring with a probability of up to 3% [2].

In this system, there is no dissipation whatsoever except during the read operation. Therefore, the product of dissipated energy (ignoring the read operation) and switching delay τ_d is exactly zero. On the other hand, the product of applied energy and switching delay is

$$h_A \tau_d = \frac{h h_A}{4\sqrt{h_A^2 + 4J^2}}. \tag{14.8}$$

We immediately see that any energy-time uncertainty limit that might have been expected is non-existent, since

$$h_A \tau_d < \frac{\hbar}{2} \qquad if\ h_A < \frac{2J}{\sqrt{\pi^2 - 1}}. \tag{14.9}$$

It should be emphasized that h_A is the energy applied to switch the inverter and does not have to be *dissipated*. Even if it were dissipated, the above equation clearly shows that there is no energy-time uncertainty limit on (dissipated) energy-delay product, contrary to the popular view espoused in [3, 4]. In a paper that is now a classic, Rolf Landauer showed that there is no lower limit imposed by the uncertainty principle on dissipated energy [5]. In fact, concrete and detailed classical models of dissipationless computation have been provided by several authors and numerous quantum mechanical models of dissipationless computation have also been forwarded starting with the early work of Benioff [6, 7, 8]. These models require no dissipated energy, but usually do require input energy to switch. What we see in the current example is that there is no energy-time uncertainty limit associated with the *applied energy* either. Furthermore, computation can proceed by applying arbitrarily small energy to initiate the process. The quantity h_A can be arbitrarily small, but the smaller we make h_A, the longer will be τ_d according to Equation (14.7). Therefore, the price we pay for being miserly with energy is long switching delay.

14.2 Can the NAND gate be switched without dissipating energy?

The reader may wonder whether the NAND gate of the previous chapter could have been switched in a similar fashion without dissipating energy. Looking at any of the Tables listing the eigenenergies and the eigenstates of the NAND gate in Chapter 13, it becomes obvious that if there are no input signals

applied to the input dots ($h_A = h_C = 0$), then the ground state of the NAND gate is

$$\psi_{NAND} = \frac{2}{\sqrt{6}}|\downarrow\uparrow\downarrow> - \frac{1}{\sqrt{6}}|\downarrow\downarrow\uparrow> - \frac{1}{\sqrt{6}}|\uparrow\downarrow\downarrow> . \tag{14.10}$$

This is an entangled state where none of the three dots has any net spin polarization[‡].

If we start with this ground state (as in the quantum inverter), apply inputs h_A and h_C, and then allow the system to evolve unitarily in accordance with Equation (14.4), will we realize NAND gate operation? In other words, will the correct outputs be produced? This is a profound question since its answer will tell us whether the NAND gate can be switched coherently without dissipating energy. The answer is actually surprising. When the inputs are [1 1], and the local magnetic fields writing input bits are in the same direction as the global magnetic field (down-spin direction), the answer is YES. With these specific inputs, the NAND gate can switch coherently without dissipation, in accordance with Equation (14.4) so that at some point in time, the 3-body wavefunction will be the unentangled state $|\downarrow\uparrow\downarrow>$ which corresponds to Fig. 13.1(a). The time required to attain this state is approximately $0.32\hbar/J$ and the requirement on the input signal strength is $h_A = h_B = h = 3J$ (see Problem 14.2).

The general wavefunction describing the state of the NAND gate is given in Equation (13.4). The amplitude of the target state $|\downarrow\uparrow\downarrow>$ is $c_3(t)$. In Fig. 14.2, we plot $|c_3(t)|^2$ as a function of time t when $h/J = 3$. Note that it reaches unity when $t \approx 0.32\hbar/J$ provided $h = 3J$. This means that the pure target state is *attainable* through coherent dynamics.

When the input bits are [1 0] or [0 1], the target states are $|\downarrow\downarrow\uparrow>$ and $|\uparrow\downarrow\downarrow>$ shown in Figs. 13.1(c) and (d), respectively. They can only be approached, but never completely attained. The amplitudes of these states are c_2 and c_5 (see Equation (13.4)). In Figs. 14.3 and 14.4, we plot the probabilities $|c_2|^2$ and $|c_5|^2$ as a function of time when $h/J = 2.03$. It can be shown that the probabilities reach a maximum when $h/J = 2.03$. Note that they never reach unity (the maximum value is about 0.97), indicating that the NAND gate CANNOT be switched coherently, without dissipation, when the inputs are [1 0] or [0 1].

The state in Fig. 13.1(b) is even less attainable, since $|c_6|^2$ is always zero. This is not surprising since the initial state ψ_{NAND} does not contain the

[‡]To find the spin polarization in any dot, algebraically add up the like spin amplitudes in that dot. For example, in dot A, the down-spin amplitude is $2/\sqrt{6}$ - $1/\sqrt{6}$ = $1/\sqrt{6}$, while the up-spin amplitude is -$1/\sqrt{6}$. The probability of dot A having up-spin is therefore 1/6 and down-spin is also 1/6. Since these probabilities are equal, the dot has no net spin polarization. If they were unequal, the dot would have had a net spin polarization. Do not think that the two probabilities should add up to 1. Since we are dealing with many-body states where the three spins act in unison and are connected, we cannot expect the probabilities in individual dots to add up to 1.

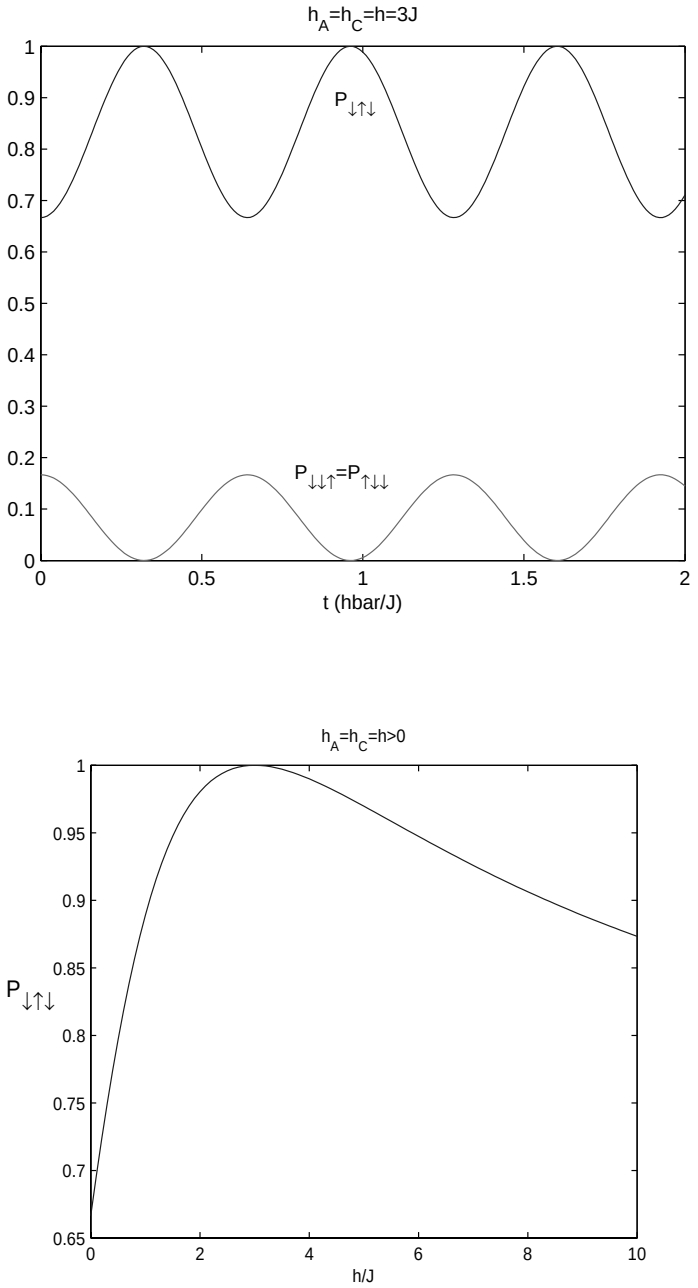


FIGURE 14.2
Probability $|c_3(t)|^2$ $(= P_{\downarrow\uparrow\downarrow})$ of the target state $|\downarrow\uparrow\downarrow>$ as a function of time when $h/J = 3$. The input bits are [1 1]. The bottom figure shows the maximum value of $|c_3(t)|^2$ as a function of h/J.

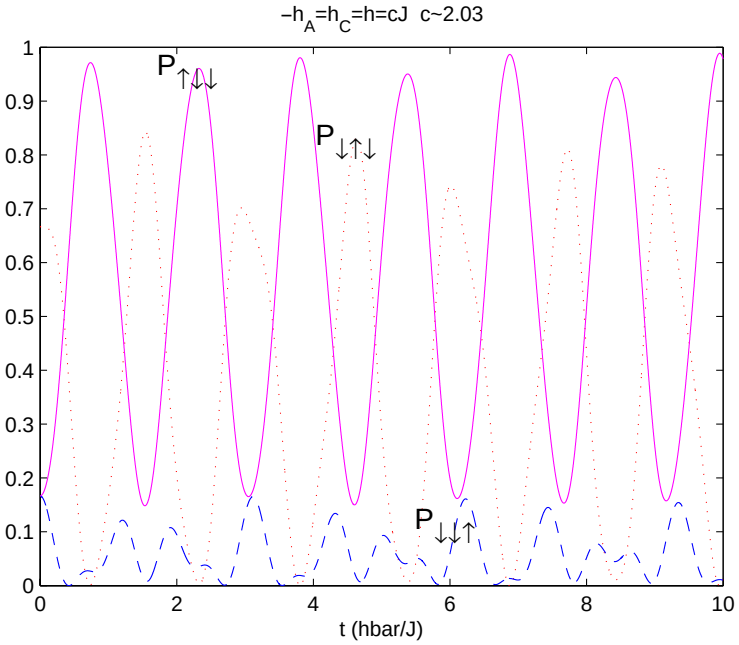


FIGURE 14.3
Probability $|c_2|^2$ $(=P_{\downarrow\downarrow\uparrow})$ as a function of time. The ratio $h/J = 2.03$. The inputs are [1 0].

target state $|\uparrow\downarrow\uparrow>$ conforming to Fig. 13.1(b), i.e., the desired state when the inputs are [0 0].

In the end, the NAND gate CANNOT be switched coherently between any two arbitrary states without dissipation. Only one entry in the truth table can be produced through coherent evolution (without dissipation) but the other three entries cannot be produced. *It is intriguing to note that the only entry in the truth table that can be produced through coherent evolution happens to be the one that is logically reversible.* When the output is 0, we know unambiguously that both inputs must have been 1, meaning that this situation is logically reversible. In this case, the gate can be switched without dissipation. In the other three cases, which are logically irreversible since the inputs cannot be inferred from the output, the gate cannot be switched coherently. This reinforces the connection between physical reversibility and logical reversibility, as originally expounded by Landauer.

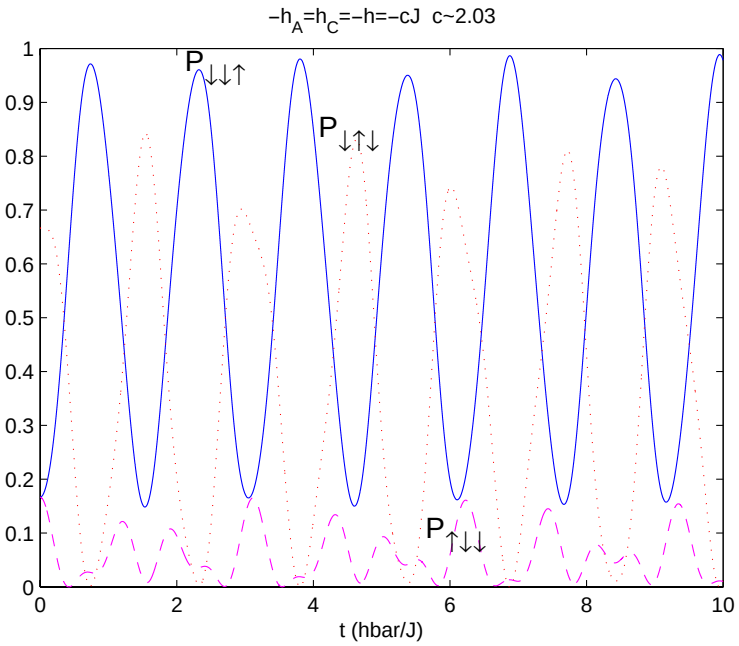


FIGURE 14.4
Probability $|c_5|^2$ $(=P_{\uparrow\downarrow\downarrow})$ as a function of time. The ratio $h/J = 2.03$. The inputs are [01].

14.3 Universal reversible gate: the Toffoli-Fredkin gate

We have seen so far that the inverter is both logically and physically reversible, while the NAND gate is both logically and physically irreversible. The inverter unfortunately is not a universal gate in the sense that any arbitrary Boolean logic circuit cannot be built with inverters alone. We therefore seek a universal reversible logic gate. Such a gate is the Toffoli-Fredkin (T-F) gate [9] that has three inputs (a, b, c) and three corresponding outputs (a', b', c').

The truth table of a T-F gate is

TABLE 14.2
Truth table of the Toffoli-Fredkin gate

a	b	c	a'	b'	c'
0	0	0	0	0	0
0	0	1	0	0	1
0	1	0	0	1	0
0	1	1	0	1	1
1	0	0	1	0	0
1	0	1	1	0	1
1	1	0	1	1	1
1	1	1	1	1	0

Looking at the above truth table, it is obvious that the input-output relationship is

$$\begin{aligned} a' &= a \\ b' &= b \\ c' &= c \oplus (a \cdot b), \end{aligned} \tag{14.11}$$

where $\oplus$ denotes the exclusive OR Boolean function and the "dot-product" denotes the Boolean AND function. The variables a and b are called control bits and the variable c is the target bit. The control bits pass through the gate without change. The target bit is flipped only if both control bits are logical 1.

We can devise a spintronic implementation of the T-F gate using the same system used for the NAND gate, namely three spins in three quantum dots (Fig. 14.5) with nearest neighbor exchange coupling. The entire array is placed in a global magnetic field which defines the spin quantization axes. In this case, the two peripheral spins are control bits a and b and the central spin is the target bit c. Assume that the global magnetic field is in the direction

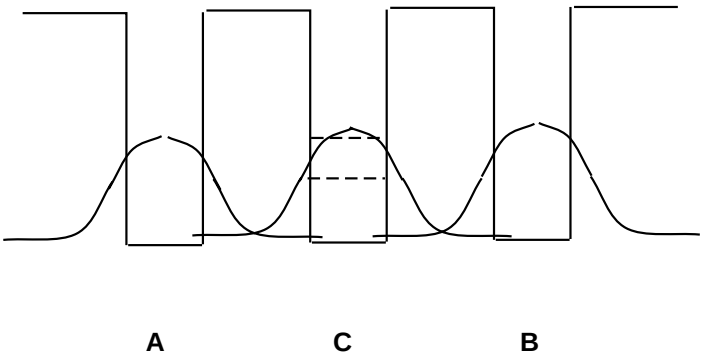


FIGURE 14.5
A Toffoli-Fredkin gate realized with three spins in quantum dots placed in a global magnetic field. Wavefunctions of nearest neighbor electrons overlap causing exchange interaction between them. Inputs are directly written in the control dots A and B with local magnetic fields that could be generated with inductors. An ac magnetic field of the right frequency flips the spin in the central dot C only when the spins in dots A and B are down, i.e., $a = b = 1$.

of "down-spin" which encodes the logic bit 1. Inputs are provided with local magnetic fields to the peripheral dots **A** and **B** to align the spins in these dots such that they conform to the intended control bits.

Consider the case when the control bits $a = b = 1$. In this case, $h_A = h_B = h$ and the eigenenergies/eigenstates are obtained from Table 13.2. The two states of interest are $|\downarrow\uparrow\downarrow>$ and $|\downarrow\downarrow\downarrow>$ since they alone correspond to the situation when the two peripheral spins are "down" representing $a = b = 1$. It is easy to show (Problem 14.5) that if $h >> J$, then the ground state is nearly the pure unentangled state $|\downarrow\uparrow\downarrow>$ while the first excited state is the pure unentangled state $|\downarrow\downarrow\downarrow>$. The energy difference between these two states is $4J - 2Z$, which is positive as long as $J > Z/2$. We will call this energy difference $\hbar\omega_{11}$ where the subscripts indicate the Boolean values of control bits a and b. Therefore,

$$\hbar\omega_{11} = 4J - 2Z. \tag{14.12}$$

Note that the above is the energy difference between states corresponding to the central spin being down and up respectively – or the target bit being 1 and 0 – when the control bits are both 1.

Similarly, when the control bits are $a = 0$, $b = 1$, i.e., $-h_A = h_B = h$, the eigenenergies/eigenstates are obtained from Table 13.4. This time, the two states of interest are $|\uparrow\uparrow\downarrow>$ and $|\uparrow\downarrow\downarrow>$ since they alone correspond to the

situation when $a = 0$, $b = 1$. Once again, we can show that when $h >> J$, the first excited state is nearly the unentangled state $|\uparrow\uparrow\downarrow>$ and the ground state is nearly the unentangled state $|\uparrow\downarrow\downarrow>$. The energy difference between them is $2Z$. Therefore,

$$\hbar\omega_{01} = 2Z. \tag{14.13}$$

Proceeding in this fashion, and always ensuring that $h >> J$, we can show that

$$\hbar\omega_{10} = 2Z \tag{14.14}$$

and

$$\hbar\omega_{00} = 4J + 2Z. \tag{14.15}$$

The purpose of all this was to show that

$$\omega_{11} \neq \omega_{00},\ \omega_{10},\ \omega_{01}. \tag{14.16}$$

14.3.1 Dynamics of the T-F gate

The T-F gate is realized if the spin in dot **C** toggles only if $a = b = 1$, and not otherwise. To achieve this, we utilize the Rabi oscillations of Chapter 4. An ac magnetic field, rotating in the x-y plane, is turned on and has an angular frequency ω_{11}. This field will flip the spin in the central dot only when $a = b = 1$, by virtue of the inequality in the last equation. This realizes the T-F gate. Note that the spin in the central dot undergoes a transition by coherently emitting or absorbing a photon from the ac magnetic field. No dissipation takes place consistent with the reversible nature of the T-F gate.

The switching time is determined by the amplitude of the ac magnetic field. It is given by

$$t_s = \frac{h}{2g\mu_B B_{ac}}, \tag{14.17}$$

where B_{ac} is the amplitude of the flux density associated with the ac magnetic field.

We can make the time t_s quite short by using quantum dot materials that have large g-factors. For a reasonable value of $B_{ac} = 0.01$ Tesla and $g = 900$, the switching time is ~ 4 picosecond. A g-factor of 900 has been calculated for certain nitrides such as $InSb_{1-x}N_x$ [10].

The ac magnetic field is turned on only for the duration t_s given by the above equation. Such a pulse is called a π-pulse. This mode of implementing reversible dynamics has been proposed by a number of authors (see, for example, [11]).

14.4 A-matrix

When dealing with reversible logic gates, a useful concept is the so-called 'A-matrix'. Consider an inverter. An inverter converts a 0 to a 1 and vice versa. Thus, the 2×2 matrix operator relating an input vector to the output vector (the vectors span the Hilbert spaces of the inputs and outputs) is given by the relation:

$$\begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix}. \\ output \; A - matrix \; input \tag{14.18}$$

The matrix operator for a logic gate is called an *A-matrix*. For an inverter, or for that matter any logically reversible gate, the A-matrix must be invertible and its determinant must not vanish. It also stands to reason that a reversible gate must have as many outputs as inputs.

We wrote the A-matrix of the basic 1-bit inverter in Equation 14.18 using the computational basis states $|0>$ and $|1>$. We can write an A-matrix for the 3-bit T-F gate using the 3-bit computational basis states $|000>$, $|001>$, $|010>$, $|011>$, $|100>$, $|101>$, $|110>$ and $|111>$ (here the basis states are possible values of $|abc>$):

$$A_{Toffoli-Fredkin} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}. \tag{14.19}$$

The A-matrix is not diagonal, but it is obviously unitary. The unitarity guarantees reversibility.

14.5 Quantum gates

The inverter and the T-F gate discussed so far are *reversible* gates and can be switched coherently in accordance of the laws of quantum mechanics. We loosely call them quantum gates, but they are not true quantum gates since they still deal with classical logic bits 0 and 1. We are now ready to explore true quantum gates.

14.5.1 The strange nature of true quantum gates: the 'square root of NOT' gate

True quantum gates are strange entities that may not have any classical analog. Let us say that we want to invert a classical bit by operating with an ordinary NOT gate. We choose to do this by operating on the input bit with two strange gates in succession that we call "square-root-of-NOT" gates denoted by $\sqrt{NOT}$. The name comes about from the fact that the successive operation of two gates can be represented mathematically as the scalar product of their A-matrices. Thus, the logic operation that we are after can be written as

$$\sqrt{NOT} \cdot \sqrt{NOT} \equiv NOT. \tag{14.20}$$

What makes this gate "quantum" (or at least non-classical) is that it is impossible to have a single-input/single-output classical binary gate that works in this fashion [12]. If the $\sqrt{NOT}$ gate were classical, it should have an output of 0 or 1 for each possible input of 0 or 1. Suppose that we define the action of a $\sqrt{NOT}$ gate as the pair of transformations

$$\begin{aligned} \sqrt{NOT}_{classical}(0) &= 0, \\ \sqrt{NOT}_{classical}(1) &= 0. \end{aligned} \tag{14.21}$$

Then, two consecutive applications of this gate will invert a 1 successfully, but not a 0, thereby violating Equation (14.20). We can try any other transformation:

$$\begin{aligned} \sqrt{NOT}_{classical}(0) &= 0. \\ \sqrt{NOT}_{classical}(1) &= 1. \end{aligned} \tag{14.22}$$

This particular transformation does not invert any bit at all. In fact, it is impossible to define $\sqrt{NOT}$ classically such that two successive operations of this gate will reproduce the behavior of a NOT gate.

The A-matrix of a $\sqrt{NOT}$ gate can be defined as

$$A_{\sqrt{NOT}} = \begin{bmatrix} \frac{1+i}{2} & \frac{1-i}{2} \\ \frac{1-i}{2} & \frac{1+i}{2} \end{bmatrix}. \tag{14.23}$$

Note that

$$\begin{bmatrix} \frac{1+i}{2} & \frac{1-i}{2} \\ \frac{1-i}{2} & \frac{1+i}{2} \end{bmatrix} \cdot \begin{bmatrix} \frac{1+i}{2} & \frac{1-i}{2} \\ \frac{1-i}{2} & \frac{1+i}{2} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \tag{14.24}$$

so that $A_{\sqrt{NOT}} \cdot A_{\sqrt{NOT}} = A_{NOT}$, as required. Moreover, the matrix $A_{\sqrt{NOT}}$ is unitary which it must be.

Note that if a $\sqrt{NOT}$ gate is fed an input data string (0,1), it produces an output data string (α, β) given by

$$\begin{aligned} \alpha &= \frac{1+i}{2}|0> + \frac{1-i}{2}|1> \\ \beta &= \frac{1-i}{2}|0> + \frac{1+i}{2}|1>, \end{aligned} \tag{14.25}$$

where the output bits are coherent superpositions of $|0>$ and $|1>$[§]. Such a bit is a 'qubit'. Since the $\sqrt{NOT}$ gate generates a qubit, it is a true quantum gate.

Consider now a different A-matrix:

$$A_{random} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}, \tag{14.26}$$

which is unitary.

If this gate is fed an input data string (0,1), it produces an output data string (α', β') given by

$$\begin{aligned} \alpha' &= \frac{1}{\sqrt{2}}|1> + \frac{1}{\sqrt{2}}|0> \\ \beta' &= \frac{1}{\sqrt{2}}|1> - \frac{1}{\sqrt{2}}|0>, \end{aligned} \tag{14.27}$$

which are also qubits.

Now a measurement of the output (either $|\alpha'>$ or $|\beta'>$) yields the answer 0 (state $|0>$) or 1 (state $|1>$) with equal probability 1/2, yielding a *perfectly random* generator of bits 0s and 1s. A single computation with a single gate yields a perfectly random bit! This is an extension of "quantum parallelism" that we will discuss later in this chapter. But here is a taste of the power of quantum computers. A classical algorithm will require more than one gate and many computational steps to generate a random number. More importantly, the number will be *never truly random*. Mathematically, there is *no function* that generates a true random number (only pseudo-random numbers can be generated using algorithms). Thus, a classical Turing machine can only generate pseudo random numbers; it cannot generate a true random number. This also shows that the classical Turing machine is not really universal since it cannot model the quantum mechanical process of generating a true random number. The gate that we have postulated here beats even the most powerful classical computer conceivable in generating random numbers.

14.6 Qubits

The "qubit" concept is at the heart of quantum computation and requires more introspection. We have seen in earlier Chapters that a qubit can be written as

$$qubit = a_0|0> + a_1|1>, \tag{14.28}$$

[§]To show this, just operate $A_{\sqrt{NOT}}$ on the vector $[0, 1]^T$.

where $|0>$ denotes the state in which the qubit has a value of 0 and $|1>$ denotes the state in which the qubit has a value of 1. The coefficients a_0 and a_1 are complex quantities whose squared magnitudes denote the probability that if a measurement is performed on the qubit, it will be found to have a value of 0 and 1, respectively. Since the measurement can yield a value of only 0 or 1, and nothing else, it follows that

$$|a_0|^2 + |a_1|^2 = 1. \tag{14.29}$$

We emphasize that the above equation *does not* imply that the qubit is sometimes in state $|0>$ with probability $|a_0|^2$, and the rest of the time in state $|1>$ with probability $|a_1|^2$. It is only after measurement that the qubit collapses to a classical bit and assumes a definite value of 0 or 1. Prior to the measurement, it *does not* have a definite value; it is both 0 and 1, *all the time*. It is therefore called a superposition state.

While one can easily understand the meaning of "superposition" from the foregoing discussion, to understand the implication of a "coherent" superposition requires more reflection. Coherence means that the phase relationship between a_0 and a_1 (being complex quantities, they have an amplitude and a phase) must be preserved. It might appear that the phase is never relevant since a_0 and a_1 seemingly determine only the probabilities which, in turn, depend only on the squared magnitudes. This is actually not true. Consider a hypothetical bit-flip operator $\hat{0}_{flip}$ whose action is to flip a bit from 0 to 1 and vice versa (a classical realization of a bit flip operator is an inverter or NOT gate). Its action on a bit can be represented as

$$\begin{aligned} \hat{0}_{flip}|0> &= 1|1> \\ \hat{0}_{flip}|1> &= 1|0> . \end{aligned} \tag{14.30}$$

Note from the above that $0>$ and $|1>$ are not eigenstates of the operator. If we want to calculate the expected value of this operator for a system described by a qubit, we will get

$$\begin{aligned} expected\ value &= < qubit|\hat{0}_{flip}|qubit> \\ &= |a_0|^2 < 0|\hat{0}_{flip}|0> + |a_1|^2 < 1|\hat{0}_{flip}|1> \\ &\quad + a_0 * a_1 < 0|\hat{0}_{flip}|1> + a_1 * a_0 < 1|\hat{0}_{flip}|0> \\ &= a_0 * a_1 + a_1 * a_0 \\ &= 2|a_0||a_1|cos\theta, \end{aligned} \tag{14.31}$$

where θ is the phase angle difference between a_0 and a_1. In deriving the last equation, we used Equations (14.29) and (14.30) and also the fact that $|0>$ and $|1>$ are orthonormal states by themselves. The fact that $|0>$ and $|1>$ are orthogonal to each other is obvious because any measurement outcome can produce either a 0 or a 1, and nothing in between. Thus the two possible outcomes are mutually exclusive, meaning that the two states are orthogonal to each other.

The above equation clearly shows that there are quantities that depend on the phases of a_0 and a_1. Thus maintaining the correct phase relations between these quantities, or "coherence", is important. Quantum computation depends critically on this coherence. This is where *spin has an advantage over charge when it comes to hosting qubits.* A qubit can be constructed with a coherent superposition of two anti-parallel spin states. This coherence time can be quite long. It has been measured as 100 nanoseconds in n-type GaAs at a temperature of 5 K [13], and it has been argued that the actual coherence time is much longer since the single spin decoherence time is longer than the ensemble averaged decoherence time that is usually measured in experiments [14]. Reference [14] claims that the single spin decoherence time will be 1 - 100 μsec. In contrast, the coherence time for charge in solids tends to saturate to about 1 nsec as the temperature of the solid is reduced to sub milliKelvins [15]. Consequently, spin-based quantum computers are much longer lived and more robust than charge based quantum computers. To put all this into perspective, let us think in terms of what happens if the coherence is lost. In that event, the quantum computer will falter and generate wrong results. An error will occur if the qubit loses coherence during a clock cycle. Therefore, the error probability is roughly $1 - exp[-T/T_2]$, where T is the clock period and T_2 is the time it takes for the qubit to decohere. In order to limit the error probability to 3%, which is the largest error probability that can be handled with error correcting codes at the time of writing this book [2], we need $T < 0.03T_2$. If $T_2 = 1$ nanosecond (charge), we need a clock frequency $1/T =$ 33 GHz, whereas if $T_2 = 1$ microsecond (spin), we need a clock frequency of 33 MHz. Alternately, with a fixed clock period of 5 GHz, the error probability with spin ($T_2 = 1$ microsecond) is 2×10^{-4}, while the error probability with charge ($T_2 = 1$ nanosecond) is 0.18. Therefore, spintronics is considerably superior to electronics when it comes to quantum computing. This realization has provided a tremendous boost to spintronics.

14.7 Superposition states

We are now ready to elucidate how superposition endows a quantum computer with immense prowess. Consider a bistable spin (a single electron placed in a magnetic field). We can use the up-spin state to encode the binary bit 1 and the down-spin state to encode the binary bit 0. Classically, this electron can then store only one bit of information (whose value can be 0 or 1) in its spin degree of freedom because we have two states (up and down) and two possible values of the binary bit. Since each electron can store only one bit, N electrons can store N binary bits classically. But now, if we can create a coherent superposition of the states of two electrons and *entangle their spins*,

the corresponding qubit will be written as

$$qubit_{2\ electrons} = a_{00}|00> + a_{01}|01> + a_{10}|10> + a_{11}|11>, \qquad (14.32)$$

where the state $|ij>$ corresponds to the first electron's spin being in state $|i>$ and the second electron's spin in state $|j>$. This system can, in principle, store 2^2 bits of information corresponding to (i) both electrons being in the down-spin state, (ii) first in down-spin state and second in up-spin, (iii) first in up-spin and second in down-spin, and (iv) both up-spin. Note that the classical case would correspond to only one of the coefficients in Equation (14.32) being non-zero. Extending this to N bistable spins, we can write the corresponding qubit as

$$qubit_{N-electron} = \sum_{x_1 x_2 ... x_N} a_{x_1 x_2 ... x_N} |x_1 x_2 ... x_N>, \qquad (14.33)$$

where the x-s can take the value 0 or 1. The above equation has 2^N terms. It means that while N bistable spins can store N binary bits classically, the same system can store 2^N classical bits quantum mechanically. Now ponder the following. If we want to store 2^{300} binary bits of information, it is impossible in the classical realm since the number 2^{300} is larger than the number of electrons in the known universe. However, in the quantum mechanical realm, just 300 electrons with entangled spins in a coherent superposition state can do the job!

Now the tempered vision. It is extremely difficult to entangle 300 qubits. So far, only a few qubits (less than 10) have been entangled. Furthermore, one might be misled to think that we can make a huge "memory" storing 2^N binary bits of information using just N entangled spins maintained in a coherent superposition state. This point of view would be *incorrect*. A reliable memory should be such that every time it is accessed, it returns the *same* data which is stored in it. For instance, if we have stored the binary string 11001, we should get this string *every time* the memory is read. But the quantum memory we are talking about does not satisfy this requirement. Every time we read the memory, the qubits collapse to classical bits and we only get N (not 2^N) bits out of it. More importantly, the values of these bits change with each measurement because of the probabilistic nature of quantum measurement. It is never the same bit string in two successive measurements! Surely, this cannot be judged a reliable memory.

14.8 Quantum parallelism

If the above is true, then how is the power of quantum computing ever harnessed? The answer is in something that has been termed "quantum parallelism". We discuss that next.

In the world of classical computers, parallelism refers to parallel (simultaneous) processing of different information in different processors. Quantum parallelism refers to simultaneous processing of different information or inputs in the *same* processor. This idea, due to Deutsch [16], refers to the notion of evaluating a function once on a superposition of all possible inputs to the function to produce a superposition of outputs. Thus, all outputs are produced in the time taken to calculate one output classically. Of course, not all of these outputs are accessible since a measurement on the superposition state of the output will produce only one output. However, it is possible to obtain certain joint properties [17] of the outputs, and that is a remarkable possibility.

Let us exemplify quantum parallelism more concretely. Consider the situation when N inputs x_1, x_2, ... x_N are provided to a computer and their functions $f(x_1)$, $f(x_2)$, ... $f(x_N)$ are to be computed. The results are then fed to another computer to calculate the functional $F(f(x_1), f(x_2), ...f(x_N))$.

With a classical computer, we will calculate $f(x_1)$, $f(x_2)$, ... $f(x_N)$ serially, one after the other. With a quantum computer, the story is different.

Prepare the initial state as a superposition of the inputs:

$$|I>= \frac{1}{\sqrt{N}}(|x_1> +|x_2> +... + |x_N>). \tag{14.34}$$

Let it evolve in time to produce the output

$$|O>= \frac{1}{\sqrt{N}}(|f(x_1)> +|f(x_2)> +... + |f(x_N)>). \tag{14.35}$$

Note that $|O>$ has been obtained in the time required to perform a single computation. Now, if $C = F(f(x_1), f(x_2), ...f(x_N))$, can be computed from $|O>$, then a quantum computer will be of great advantage. This is an example where "quantum parallelism" can be used to speed up the computation tremendously.

There are two questions now. First, can C be computed from a knowledge of the superposition of various $f(x_i)$-s and not the individual $f(x_i)$-s? The answer is "yes", but for a small class of problems. These are called the Deutsch-Josza class of problems which can benefit from quantum parallelism. Second, can C be computed correctly with unit probability? The answer is "no". C cannot be computed with unit probability. However, if the first answer is wrong (hopefully, the computing entity can differentiate right from wrong answers), then the experiment or computation is repeated until the right answer is obtained. The probability of getting the right answer within k iterations is $(1-p^k)$ where p is the probability of getting the wrong answer in any iteration. The mean number of times the experiment should be repeated is $N^2 - 2N - 2$.

The following is an example of Deutsch-Josza class of problems. For integer $0 \leq x \leq 2L$, given that the function $f_k(x) \in$ [0,1] has one of two

properties – (i) either $f_k(x)$ is independent of x, or (ii) one half of the numbers $f_k(0), f_k(1), ...f_k(2L-1)$ are zero – determine which type the function belongs to using the fewest computational steps.

The most efficient classical computer will require $L+1$ evaluations, whereas according to Deutsch and Josza, a quantum computer can solve this problem with just two iterations.

14.9 Universal quantum gates

The important question that concerns engineers and applied scientists is whether there exists a *universal* quantum gate. If it exists, we will focus on how to implement and synthesize this gadget, since it will ultimately lead to a quantum computer. The first universal quantum gate that was postulated is a 3-qubit gate due to Deutsch. It is fashioned after the universal classical gate of Toffoli and Fredkin. Using the computational basis states $|000>$, $|001>$, $|010>$, $|011>$, $|100>$, $|101>$, $|110>$, and $|111>$, the A-matrix of the Deutsch gate is given by

$$A_{Deutsch} = \begin{bmatrix} 1&0&0&0&0&0&0&0 \\ 0&1&0&0&0&0&0&0 \\ 0&0&1&0&0&0&0&0 \\ 0&0&0&1&0&0&0&0 \\ 0&0&0&0&1&0&0&0 \\ 0&0&0&0&0&1&0&0 \\ 0&0&0&0&0&0&A&B \\ 0&0&0&0&0&0&B&A \end{bmatrix}, \tag{14.36}$$

where

$$A = ie^{\frac{i\pi\alpha}{2}}(1+e^{i\pi\alpha}) \; and \; B = ie^{\frac{i\pi\alpha}{2}}(1-e^{i\pi\alpha}) \tag{14.37}$$

and α may be an irrational number.

It is easy to check that $A_{Deutsch}$ is unitary. The T-F gate can be synthesized by connecting a number of Deutsch gates in series. The number of Deutsch gates required for this is N where N must satisfy the relation $[A_{Deutsch}]^N = A_{Toffoli-Fredkin}$. Noting that

$$\begin{bmatrix} ie^{\frac{i\pi\alpha}{2}}(1+e^{i\pi\alpha}) & ie^{\frac{i\pi\alpha}{2}}(1-e^{i\pi\alpha}) \\ ie^{\frac{i\pi\alpha}{2}}(1-e^{i\pi\alpha}) & ie^{\frac{i\pi\alpha}{2}}(1+e^{i\pi\alpha}) \end{bmatrix}^N = i^N e^{\frac{iN\pi\alpha}{2}} \begin{bmatrix} ie^{\frac{i\pi\alpha}{2}}(1+e^{i\pi\alpha}) & ie^{\frac{i\pi\alpha}{2}}(1-e^{i\pi\alpha}) \\ ie^{\frac{i\pi\alpha}{2}}(1-e^{i\pi\alpha}) & ie^{\frac{i\pi\alpha}{2}}(1+e^{i\pi\alpha}) \end{bmatrix} \tag{14.38}$$

the quantity $e^{iN\pi\alpha}$ can be made arbitrarily close to any complex number of unit norm by changing N. Choosing N such that $e^{iN\pi\alpha}$ = -1 yields the Toffoli-Fredkin gate [18].

14.9.1 2-qubit universal quantum gates

Later, it was shown that there exist 2-qubit quantum gates which are universal. The proof of universality of these gates was given by DiVincenzo [19] using Lie algebra; however, it can also be shown that the Deutsch gate can be realized with these 2-qubit gates. This, in itself, is sufficient proof of universality. The Deutsch gate is not the most primitive universal quantum gate; the DiVincenzo gate is more primitive. Later, it was shown by Seth Lloyd that almost any 2-qubit gate is universal [20].

The computational basis of the 2-terminal (2 inputs and 2 outputs) universal quantum gate can be chosen as $|00>$, $|01>$, $|10>$, $|11>$. In this basis, the A-matrix of the universal 2-qubit gate is

$$A_{DiVincenzo} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & A & B \\ 0 & 0 & B & A \end{bmatrix}, \tag{14.39}$$

where A and B have been given by Equation 14.37.

14.10 A 2-qubit "spintronic" universal quantum gate

Today, most experimental effort in realizing semiconductor solid-state versions of universal 2-qubit gates is focused on the realization of a quantum controlled NOT or XOR gate. These are the two most popular gates and both are "universal". In these gates, one of the two qubits is called the target qubit and the other is the control qubit. The basic operation of the controlled NOT gate has two ingredients: (i) arbitrary rotation of the target qubit, and (ii) controlled rotation of the target qubit through specified angles if and only if the control qubit is in a certain state. The controlled XOR gate has similar dynamics: (i) arbitrary rotation of the target qubit, and (ii) a so-called "square-root-of-swap" operation where the quantum states of the target and control qubits are half-way interchanged [21].

Numerous authors have proposed various spintronic embodiments of the 2-qubit universal quantum gate. It is not possible to discuss all, or even a large subset of them, in a textbook. Therefore, we take a historic perspective and discuss some of the early ideas for universal quantum gate.

14.10.1 The silicon quantum computer based on nuclear spins

The first comprehensive solid state scalable realization of a universal 2-qubit spintronic quantum gate is probably the one due to Kane [22]. Qubits are encoded in the nuclear spin orientations of ^{31}P dopant atoms in silicon. These

nuclear spin orientations are coherent superpositions of two anti-parallel polarizations. The coherence is long lived since nuclear spins have very long coherence times, possibly several milliseconds at 4.2 K temperature. Therefore, the qubits are robust.

To select a qubit for rotation, the spin-splitting energy of the target nucleus is first altered (using hyperfine interaction) by changing the wave function of the electron bound to that nucleus. This is achieved by applying a potential via a lithographically defined gate (called an "A-gate") placed precisely on top of that nucleus. The gate potential attracts or repels the electron bound to that nucleus thereby changing its wave function, and hence the nucleus' spin-splitting energy via hyperfine interaction. The spin-splitting energy is made resonant with a global ac magnetic flux density of amplitude B_{ac} for specific duration (Rabi oscillation). If the duration is τ, then the angle by which the spin is rotated is

$$\theta(\tau) = \frac{1}{\hbar} \int_0^\tau \vec{\mu}_B \cdot \vec{B}_{ac} dt. \tag{14.40}$$

This realizes single qubit rotation.

The 2-qubit operation is achieved by yet another set of lithographically defined gates (called "J gates") which raise or lower a potential barrier between two neighboring nuclei's electrons (one nucleus is the "target" and the other is the "control" qubit). As the barrier is lowered, the wave functions of the two electrons bound to those nuclei overlap and the resulting exchange interaction lowers the energy of the singlet state with respect to the triplet state (as long as there is no magnetic field to induce a Zeeman splitting that exceeds the exchange splitting). As a result, the spin-splitting energy in any nucleus depends on the spin state of its neighbor. Thus, the rotation of the target qubit by the global ac magnetic field can be made conditional on the spin state of the control qubit. This realizes the quantum controlled NOT operation. When the gate operations are over, the nuclear spin is "read" by transducing the nuclear spin to electron spin, and then reading that electron spin with spin transistors or other devices that are capable of spin-to-charge transduction, and thus allowing a measurement of the spin.

A closely related proposal that hosts qubits in electronic spin, rather than nuclear spin, has been forwarded by Vrijen et al. [23]. This model eliminates the need to transduce the qubit between nuclear and electron spin. Instead of using hyperfine interaction to change the spin-splitting energy in a target qubit, they utilize a Si/SiGe heterostructure. The ^{31}P dopant atom has to be placed exactly at the interface of these two materials and a lithographically delineated gate has to be aligned exactly on top of this buried dopant atom. The gate potential pushes the electron wavefunction into either the SiGe side, or the Si side. Since these two materials have different g-factors, the spin-splitting energy in a constant magnetic field can be varied by the gate potential. This eliminates the need for hyperfine interaction (which is weak)

to effect single qubit rotations. The downside is that extreme precision is required to place the buried dopant atom exactly underneath the gate (lateral precision) and also exactly at the interface of Si and SiGe (vertical precision). The lateral alignment may be achieved by ion-implanting the dopant through a mask [24], but the vertical alignment is much more difficult. One has to virtually eliminate "straggle" in ion implantation, and that is extremely difficult.

14.10.2 Quantum dot-based spintronic model of universal quantum gate

The idea of using the spin of a single electron in a quantum dot to serve as a vehicle for qubits has been around since at least 1996 [1, 25]. This has spawned many proposals for realizing universal quantum gates using the spin of a single electron in a quantum dot. We outline one such proposal [26].

Consider a penta-layered semiconductor quantum dot defined by split-gates on a heterostructure consisting of ferromagnet-insulator-semiconductor-insulator-ferromagnet layers as shown in Fig. 14.6. We also show the conduction band diagram in the direction perpendicular to the heterointerfaces. The ferromagnets are magnetized and generate a permanent magnetic field over the semiconductor which spin-splits the quantum dot energy levels in the semiconductor because of Zeeman interaction. There may also be additional spin-splitting due to spin-orbit interaction. The Fermi level, which is determined by the metallic ferromagnetic contacts, is initially below the lower spin-split level, so that the quantum dot is depleted of all electrons (we assume very low temperature for this discussion).

Next, a positive potential is applied to the split-gates which effectively makes the quantum dot slightly larger and pulls the lower Zeeman spin-split level below the Fermi level. A spin polarized electron (whose spin is the majority spin in the ferromagnet) now tunnels into the semiconductor dot from the contacts. Only one electron can enter provided its spin is the majority spin in the ferromagnet because of Pauli exclusion principle. Therefore we have "written" and initialized a qubit.

The insulating layers provide a confining potential for the electron in the semiconductor dot. They also facilitate coherent spin injection from the ferromagnet because they increase the spin injection efficiency [27].

To select a qubit for rotation, we apply a differential potential between the two arms of the split-gate to cause a structural inversion asymmetry and an ensuing Rashba interaction. This modifies the total spin-splitting energy in the quantum dot (see Chapter 7). We apply just enough differential potential to make the total spin-splitting energy resonant with a global ac magnetic field. We hold the resonance for a finite duration τ. This rotates the spin in accordance with Equation (14.40). By controlling the time duration of the resonance, i.e., the quantity τ in Equation (14.40), we can rotate the spin through arbitrary angles θ. This satisfies the first requirement for a universal

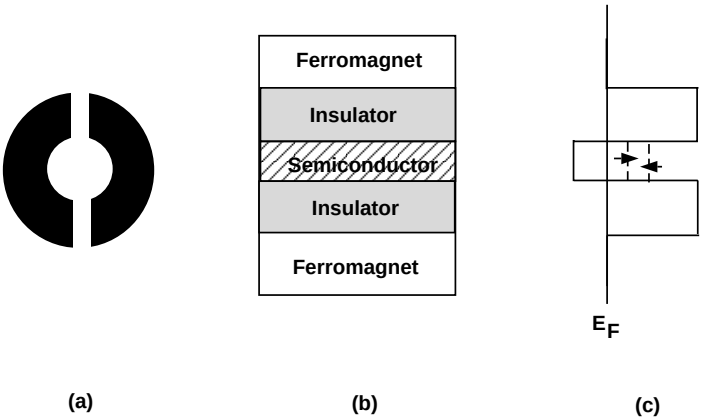


FIGURE 14.6
(a) Top view of the quantum dot defined by ring shaped split-gates. (b) A pentalayered heterostructure. The electron is confined in the semiconductor layer. (c) The conduction band energy diagram in the direction perpendicular to the heterointerfaces. The Fermi level is determined by the metallic ferromagnets. The broken lines show the initial spin-split energy levels in the semiconductor dot.

quantum gate, namely arbitrary single qubit rotation.

The second and last ingredient is the 2-qubit controlled NOT or square-root-of-swap operation. For this purpose, we apply a positive potential to the barrier separating two adjacent quantum dots that lowers the barrier and makes the wavefunctions in these two dots overlap. That generates an exchange interaction. If the exchange interaction is allowed to take place for a duration $h/(4J)$, where J is the exchange interaction strength, we will have realized the 'square-root-of-swap' operation [21]. The controlled NOT operation can also be realized since the spin-splitting energy in one dot (the difference between the up-spin and the down-spin energies) definitely depends on the spin state in its neighbor because of exchange interaction (recall the quantum inverter). Because of this, we can make the spin-splitting energy in the target dot resonant with a global magnetic field only if its neighbor's spin is up or down. Simply put, we can make the spin rotation in the target dot conditional on the spin state in the control dot. This realizes the controlled NOT operation, which is the second (and final) ingredient of a universal 2-qubit quantum gate.

14.11 Conclusion

In conclusion, we have provided the reader with a brief glimpse of spin-based quantum computing. We have omitted any discussion of opto-spintronic implementations of quantum gates since they are more complex and require introduction of ideas and topics that are not in this textbook. These implementations are however (see, for example, [28]) important and probably more numerous than the purely spintronic implementations. Spin based quantum computing is an area that is progressing very rapidly at this time. The Los Alamos e-print servers (www.arXiv.org.cond-mat/ and www.arXiv.org.quant-ph/) archive important literature in this field which is updated daily. The reader is strongly encouraged to consult this resource for the ever expanding knowledge base on this subject.

14.12 Problems

- **Problem 14.1**

  Using Equation (14.6), show that $|c_2(\tau_d)| = 1$ where τ_d is given by Equation (14.7) and the following condition holds: $h_A = 2J$.

- **Problem 14.2**

  Consider a NAND gate. Using Equation (14.4) and noting that the Hamiltonian H is given by Equation (13.3), show that $\psi(t_d) = |\downarrow\uparrow\downarrow>$, where $t_d \approx 0.32\hbar/J$, provided $h = 3J$. You can use software tools like MATLAB for symbolic manipulation.

- **Problem 14.3**

  When the input bits to a NAND gate are [0 0], show that the amplitude of the state $|\downarrow\uparrow\downarrow>$ reaches zero at time $0.32\ \hbar/J$ when $h = 3J$.

- **Problem 14.4**

  When the input bits to a NAND gate are [0 1] or [1 0], show that the amplitudes of the target state $|\uparrow\downarrow\downarrow>$ or $|\downarrow\downarrow\uparrow>$ reaches a maximum when $h \approx 2J$.

- **Problem 14.5**

  Show that when $h >> J$, the ground state in Table 13.2 is nearly the pure unentangled state $|\downarrow\uparrow\downarrow>$. Hint: Evaluate the coefficients of the three different constituents of the ground state and show that the coefficient of the $|\downarrow\uparrow\downarrow>$ constituent reaches nearly unity when $h >> J$.

14.13 References

[1] S. Bandyopadhyay and V. P. Roychowdhury, "Switching in a reversible spin logic gate", Proc. Superlat. Microstruct. Conf., Liege, Belgium, July 1996. (Superlat. Microstruct., **22**, 411 (1997)).

[2] E. Knill, "Quantum computing with realistically noisy devices", Nature (London), **434**, 39 (2005).

[3] R. T. Bate in *VLSI Microstructure Science and Engineering*, Vol. 4, Ed. N. G. Einspruch, (Academic Press, New York, 1981).

[4] C. Mead and L. Conway, *Introduction to VLSI Systems*, (Addison-Wesley, Reading, MA, 1980).

[5] R. Landauer, "Uncertainty principle and minimal energy dissipation in the computer", Intl. J. Theor. Phys., **21**, 283 (1982).

[6] P. Benioff, "The computer as a physical system- A microscopic quantum mechanical Hamiltonian model of computers as represented by Turing machines", J. Stat. Phys., **22**, 563 (1980).

[7] P. Benioff, "Quantum mechanical models of Turing machines that dissipate no energy", Phys. Rev. Lett., **48**, 1581 (1982).

[8] P. Benioff, "Quantum mechanical Hamiltonian models of discrete processes that erase their own histories: Applications to Turing machines", Intl. J. Theor. Phys., **21**, 177 (1982).

[9] E. Fredkin and T. Toffoli, "Conservative logic", Intl. J. Theor. Phys., **21**, 219 (1982).

[10] X. W. Zhang, W. J. Fan, S. S. Li and J. B. Xia, "Giant and zero electron g-factors of dilute nitride semiconductor nanowires ", Appl. Phys. Lett., **90**, 193111 (2007).

[11] S. Lloyd, "A potentially realizable quantum computer ", Science, **261**, 1569 (1993).

[12] C. P. Williams and S. H. Clearwater, *Explorations in Quantum Computing*, (Springer-Verlag, New York, 1998).

[13] J. M. Kikkawa and D. D. Awschalom, "Resonant spin amplification in n-type GaAs ", Phys. Rev. Lett., **80**, 4313 (1998).

[14] R. deSousa and S. Das Sarma, "Electron spin coherence in semiconductors: Considerations for a spin-based solid-state quantum computer architecture", Phys. Rev. B, **67**, 033301 (2003).

[15] P. Mohanty, E. M. Q. Jariwalla and R. A. Webb, "Intrinsic decoherence in mesoscopic systems", Phys. Rev. Lett, **78**, 3366 (1997).

[16] D. Deutsch, "Quantum theory, the Church-Turing principle and the universal quantum computer ", Proc. Royal Soc. London A, **400**, 97 (1985).

[17] R. Jozsa, "Characterizing classes of functions computable by quantum parallelism", Proc. Royal Soc. London A, **435**, 563 (1991).

[18] V. P. Roychowdhury, private communication.

[19] D. P. DiVincenzo, "2 bit gates are universal for quantum computation", Phys. Rev. A., **51**, 1015 (1995).

[20] S. Lloyd, "Almost any quantum logic gate is universal ", Phys. Rev. Lett., **75**, 346 (1995).

[21] G. Burkard, D. Loss and D. P. DiVincenzo, "Coupled quantum dots as quantum gates", Phys. Rev. B, **59**, 2070 (1999).

[22] B. E. Kane, "A silicon based nuclear spin quantum computer ", Nature (London), **393**, 133 (1998).

[23] R. Vrijen, et al., "Electron spin resonance computers for quantum computing in silicon-germanium heterostructures", Phys. Rev. A, **62**, 12306 (2000).

[24] R. P. McKinnon, et al., "Nanofabrication processes for single ion implantation of silicon quantum computer devices", Smart Materials and Structures, **11**, 735 (2002).

[25] A. M. Bychkov, L. A. Openov and I. A. Semenihin, "Single electron computing without dissipation", JETP Lett., **66**, 298 (1997).

[26] S. Bandyopadhyay, "Self assembled nanoelectronic quantum computer based on the Rashba effect in quantum dots", Phys. Rev. B, **61**, 13813 (2000).

[27] E. I. Rashba, "Theory of electrical spin injection: Tunnel contacts as a solution of the conductivity mismatch problem", Phys. Rev. B, **62**, R16267 (2000).

[28] T. Calarco, A. Datta, P. Fedichev, E. Pazy and P. Zoller, "Spin based all optical quantum computation with quantum dots: Understanding and suppressing decoherence", Phys. Rev. A., **68**, 012310 (2003).

15

A Quantum Mechanics Primer

This chapter should be treated as an appendix. It is included for completeness and covers certain fundamentals of Quantum Mechanics that will aid the understanding of topics discussed in this textbook. The reader is urged to refer to many of the good introductory textbooks on the subject to learn or revisit the fundamental concepts [1, 2, 3, 4].

First, we briefly discuss three key ingredients of quantum mechanics, i.e., the blackbody radiation spectrum, the concept of the photon and de Broglie wavelength, and the related wave-particle duality concept. Next, we briefly restate the four major postulates of Quantum Mechanics used in the so-called *orthodox* or Copenhagen interpretation of quantum mechanics that is in vogue. We continue on to describe a generalized form of the Heisenberg Uncertainty Principle which is illustrated with a few familiar bound state problems (particle-in-a-box and one-dimensional harmonic oscillator) and a few spin related problems.

15.1 Blackbody radiation and quantization of electromagnetic energy

15.1.1 Blackbody radiation

A watershed event in quantum physics was Max Planck's derivation of the correct expression for the experimentally measured energy per unit volume per frequency interval (or energy spectral distribution) of the blackbody radiation

$$u = \frac{8\pi h\nu^3}{c^3}\frac{1}{e^{\frac{h\nu}{k_B T}} - 1}, \tag{15.1}$$

where h is Planck's constant, ν is the frequency of the electromagnetic radiation in the cavity, c is the speed of light in vacuum, k_B is Boltzmann's constant, and T is the temperature inside the cavity in Kelvin. Planck was able to derive this important relation by assuming that the energy exchange between the electromagnetic waves inside a cavity and its walls occurs via emission and absorption of discrete quanta of energy. As a result, the energy

of the electromagnetic radiation with frequency ν inside the blackbody cavity exists only in multiples of $h\nu$.

Exercises:

- Show that the maximum of the energy spectral density u in Equation (15.1) occurs when
$$3 - 3exp(-x) = x, \tag{15.2}$$
where $x = h\nu/k_BT$.

- Solve the equation above numerically and show that the maximum occurs for x around 2.82, i.e., at a frequency ν_{max} given by
$$h\nu_{max}/k_BT \sim 2.82. \tag{15.3}$$
This last relation is referred to as *Wien's displacement law*.

- According to the previous exercise, the peak in the energy spectral distribution curve shifts upward in frequency as the temperature increases. What is the equilibrium temperature inside a blackbody cavity when the maximum in the energy spectral distribution is to be located at wavelengths 1,000 $\mathring{A}$, 5,000 $\mathring{A}$, and 10 μm, which correspond to the ultraviolet, visible, and mid-infrared region of the electromagnetic spectrum, respectively?

15.2 The concept of the photon

Following Planck's seminal work, Einstein postulated that electromagnetic radiation consists of particle-like discrete bundles of energy called *photons*. Furthermore, each photon has an energy E that depends only on its frequency ν according to the relation
$$E = h\nu = \frac{hc}{\lambda}, \tag{15.4}$$
where the last equality is simply obtained from the connection between wavelength and frequency for an electromagnetic wave propagating in vacuum.

The concept of the photon was put on an even more solid ground as Einstein successfully refined the concept in 1905 to give the first correct explanation of the photoelectric effect, one of the major puzzles in physics in the early 1990s. Incredibly, during the same year, Einstein introduced to the world of physics his special theory of relativity in which all photons (in vacuum) travel at the speed of light, irrespective of their frequency and the frame of reference

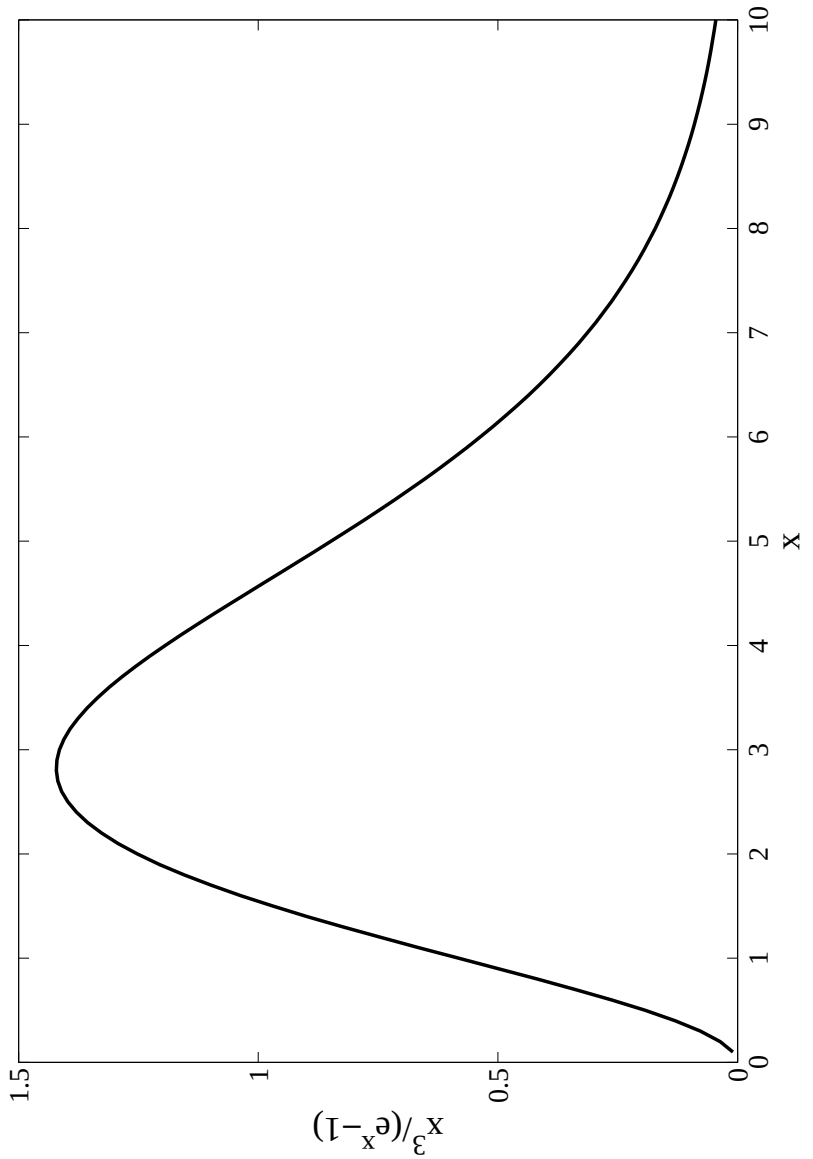


FIGURE 15.1
Plot of u, the blackbody radiation energy per unit frequency interval per unit volume (in units of $\frac{8\pi}{h^2c^3}(k_BT)^3$); x is defined as $h\nu/k_BT$.

used (for frames related by the so-called Lorentz transformation). According to the special theory of relativity, photons have a zero rest mass and their energy is entirely kinetic. The relativistic energy momentum relationship for a photon is given by

$$E = pc, \tag{15.5}$$

i.e., a photon in vacuum has a momentum simply related to its wavelength according to the formula

$$p = \frac{E}{c} = \frac{h\nu}{c} = \frac{h}{\lambda}. \tag{15.6}$$

Thus, according to the quantum point of view, a beam of electromagnetic radiation in free space can be thought as a beam of photons traveling at the speed of light c. The intensity of the beam is proportional to the number of photons crossing a unit area per unit time. For a monochromatic beam, the intensity I is given by

$$I = Nh\nu/(At), \tag{15.7}$$

where N is the number of photons falling on an area A during the time interval t.

For convenience in calculations, the following expressions in non-standard units can be used

$$h = 4.136 \times 10^{-15} eVs, \tag{15.8}$$

$$hc = 12.4 keV\mathring{A}, \tag{15.9}$$

and

$$E = \frac{hc}{\lambda} = \frac{1.24eV}{\lambda(\mu m)}, \tag{15.10}$$

where the wavelength must be expressed in μm in the last equality to get the energy in eV.

Exercises:

- What is the wavelength and frequency associated with a photon with 1 GeV energy?

  Answer: $\nu = 2.42 \times 10^{23}$ Hz; $\lambda = 1.24 \times 10^{-5}\mathring{A} = 1.24$ fm.

- Determine the photon flux associated with a beam of monochromatic light at 5,000 $\mathring{A}$ and with an intensity of 100 W/m^2.

  Solution

  $$E = h\nu = \frac{hc}{\lambda} = \frac{(6.63 \times 10^{-34} Js)(3 \times 10^{-8} m/s)}{5 \times 10^{-7} m} = \frac{6.63 \times 3}{5} 10^{-19} J. \tag{15.11}$$

  Hence

  $$N/(At) = \frac{I}{h\nu} = 25 \times 10^{19} \frac{photons}{sec.m^2}. \tag{15.12}$$

15.3 Wave-particle duality and the De Broglie wavelength

Planck's and Einstein's work established beyond doubt that electromagnetic radiation must have a particle nature in order to explain certain experimental observations, such as the photoelectric effect [5]. This is reminiscent of Newton's corpuscular theory of light, although the two are not connected. However, interference and diffraction experiments showed that electromagnetic radiation can also behave like a wave. Hence, electromagnetic radiation has a dual character; it exhibits a wave-particle duality. In certain circumstances it behaves like a wave, while in other circumstances, it behaves as a beam of particles.

Bohr postulated the orthodox *complementarity principle* that made the particle character and the wave character mutually exclusive [6]. It asserted that in situations where the particle character is manifested, the wave character is completely suppressed and vice versa. This orthodox principle has since been disproved by a remarkable set of experiments. Ghose, Home and Agarwal [7] had proposed a biprism experiment schematically depicted in Fig. 15.2. A single photon source emits a single photon which is split into orthogonal states ψ_r and ψ_t by a 50:50 beam splitter. They are detected by two photon detectors D_r and D_t. If the photon behaves truly as a "particle", then it should be detected at either D_r or D_t (but never at both) since a particle cannot traverse two paths simultaneously. That is, there should be perfect anti-coincidence between D_r and D_t, or, in other words, *either* D_r *or* D_r will click but *both* will never click simultaneously in between the arrival of two successive photons. The clever twist in this experiment, motivated by an experiment performed in the 19th century by Jagadish Chandra Bose [8], is the placement of the biprism with a small tunneling gap in the path of the transmitted photon. If D_t clicks *and* D_r *does not*, then we have made a "which path" (*welcher weg*) determination (the particle took the path of transmission as opposed to reflection) and a sharp *particle nature* is demonstrated [9]. Yet, to arrive at D_t, the particle must have tunneled through the biprism and tunneling is a sharp *wave* attribute. Therefore, this gedanken experiment presents a conundrum: the photon is behaving as both a particle and a wave in the *same experiment*, disproving the Bohr Complementarity Principle.

This experiment was actually conducted by Mizobuchi and Ohtake [10], using a single photon source. Perfect anti-coincidence was found between D_t and D_r demonstrating the particle nature. Yet, the very fact that D_t ever clicked required tunneling and hence the existence of a wave nature. It was claimed that in this experiment, a photon was behaving both sharply as a particle and as a wave in violation of orthodox Bohr's Complementarity Principle. A slight modification of this experiment has been proposed by Rangwala and Roy [11] where interference is used instead of tunneling to showcase the wave-like behavior. They claimed that quantum mechanics does not prohibit the

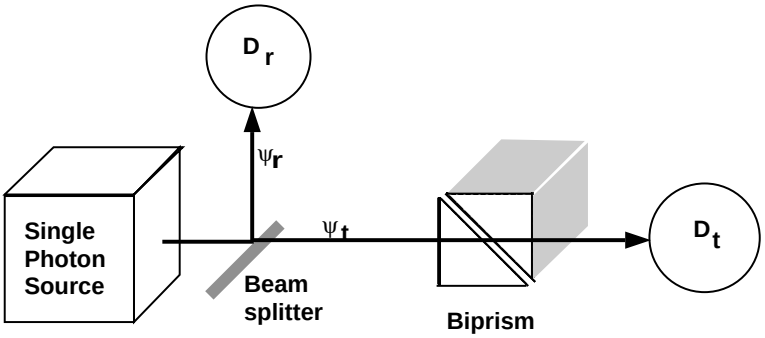


FIGURE 15.2
An experiment to disprove the orthodox Bohr's complementarity principle.

demonstration of simultaneous wave and particle behavior; rather, it prohibits their simultaneity only when the wave and particle attributes are "complementary" [12] in the sense that projection operators associated with them do not *commute*. Their view is that in the experiments of refs. [7, 10, 11], the wave and particle behavior supposedly are not truly complementary and hence not subject to the restrictions of the orthodox Complementarity Principle.

A somewhat tempered version of the Complementarity Principle is the *duality principle* due to Englert [13] which states that a quantum system can simultaneously exhibit wave and particle behavior, but sharpening of the wave character blurs the particle character and vice versa. This is an extension of earlier work by Wootters and Zurek [14], Greenberger and Yasin [15], and Mandel [16]. In fact, Englert derives an inequality

$$P^2 + V^2 \leq 1, \tag{15.13}$$

where P is a measure of the particle attribute and V is a measure of the wave attribute. Equation (15.13) immediately shows that stronger wave or stronger particle behavior can be manifested only at the expense of each other.

Returning to the wave-particle duality, Louis de Broglie made another breakthrough in 1924 by postulating that if electromagnetic radiation could have a dual wave-particle character, then material objects will also behave like waves. In other words, de Broglie posited that if a particle passes through a slit whose width is comparable to the wavelength associated with it, then it should undergo diffraction just as photons do in a single-slit experiment.

Louis de Broglie conjectured that the wavelength associated with a material body would satisfy the same relation as photons, a conjecture which Einstein himself considered as brilliant. Therefore, a material body with a momentum

p will have a wavelength given by

$$\lambda = h/p. \tag{15.14}$$

However, there is one important difference between photons and massive objects in the way their wave and particle properties are related. Because $\lambda\nu = c$ for a photon, only one relation is required to get both wavelength and frequency from a photon's particle properties of energy and momentum. A massive object, on the other hand, requires separate equations to calculate its wavelength and frequency. For a relativistic particle, the total energy E (kinetic + rest mass) is related to momentum p according to the relation

$$E = \sqrt{(pc)^2 + {E_0}^2}, \tag{15.15}$$

where $E_0 = m_0c^2$ and m_0 is the rest mass of the particle. We use this relation in Chapter 2 to ultimately derive the Dirac Equation.

The total energy E is also related to the kinetic energy K via the relation

$$E = K + m_0c^2. \tag{15.16}$$

Therefore, according to de Broglie, the wavelength associated with a relativistic particle must be determined as follows

$$\lambda_r = \frac{hc}{pc} = \frac{hc}{\sqrt{E^2 - {E_o}^2}} = \frac{hc}{\sqrt{(K + E_0)^2 - {E_0}^2}}. \tag{15.17}$$

Exercises:

- What is the accelerating potential needed to give an electron a de Broglie wavelength of 5 $\mathring{A}$, the size of the typical unit cell in many of the commonly studied semiconductors?

  Solution

  For an electron (with a initial zero kinetic energy) accelerated by a potential difference ΔV, its kinetic energy K becomes

  $$K = e\Delta V = \frac{p^2}{2m_0} = \frac{1}{2m_0}(\frac{h}{\lambda})^2. \tag{15.18}$$

  For $\lambda = 5\mathring{A}$ and m_0 equal to free electron mass, the accelerating potential $\Delta V \sim 6$ V.

- Calculate the de Broglie wavelength of an electron with a kinetic energy equal to the longitudinal optical phonon energy ($\hbar\omega_o = 36meV$)

in GaAs. The effective mass of the electron is assumed to be $m^* = 0.067m_o$, m_o being the free electron mass.

Solution

For an electron with kinetic energy $\frac{p^2}{2m^*} = \hbar\omega_o = 36meV$ and $m^* = 0.067m_o$, the de Broglie wavelength is

$$\lambda = \frac{h}{p} = \frac{h}{\sqrt{2m^*\hbar\omega_o}} \simeq 250\mathring{A}. \tag{15.19}$$

- The non-relativistic limit of expression (15.17) is given by

$$\lambda_{nr} = \frac{h}{p} = \frac{h}{\sqrt{2m_0K}}. \tag{15.20}$$

For what value of the ratio K/E_0 will the ratio of the non-relativistic and relativistic expressions for the de Broglie wavelength differ by 10%? What kinetic energy does this amount to for an electron? (The rest mass of an electron is 511 keV.)

Solution

We get

$$\frac{\lambda_{nr}}{\lambda_r} = \sqrt{1 + \frac{1}{2}(\frac{K}{E_0})}. \tag{15.21}$$

This ratio is equal to 1.1 when $\frac{K}{E_0} = 0.42$. This corresponds to a kinetic energy of 215 keV for an electron.

15.4 Postulates of quantum mechanics

This section closely follows the treatment of Nielsen and Chuang in their book titled *Quantum Computation and Quantum Information* [17]. A recap of fundamental concepts such as vector space, inner product, outer product, Hilbert space, linear operators, matrices, eigenvectors, adjoint, unitary and Hermitian operators, tensor products, operator functions, commutators and anti-commutators is given in chapter 2, section 2.1 of ref. [17].

Postulate 1: Associated with any isolated physical system is a complex vector space with inner product (Hilbert space) known as the *state space* of the system. The system is completely described by its *state vector*, which is a unit vector in the state space.

In many chapters of this textbook, we are mostly concerned with a single qubit and will work in the Hilbert space $\mathcal{C}^2$. Any state vector in that Hilbert

space is of the general form $|\psi>= a|0> +b|1>$, where a and b are complex numbers and the kets $|0>$ and $|1>$ are the 2×1 column vectors

$$|0>= \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \tag{15.22}$$

and

$$|1>= \begin{bmatrix} 0 \\ 1 \end{bmatrix}. \tag{15.23}$$

Since $|\psi>$ is a unit vector in that Hilbert space, we must have $<\psi|\psi>= 1$, i.e., $|a|^2 + |b|^2 = 1$, which is referred to as the normalization condition.

The extension of Postulate 1 to a multiparticle system (referred to as **Postulate 4** of quantum-mechanics) is based on the concept of tensor product. A rigorous treatment of the exchange interaction between two spins (recall Chapter 9) representing two qubits (recall the square-root-of-swap operation in Chapter 14) would have required the use of tensor products of the two qubits.

Postulate 2: The evolution of a closed quantum system is described by a unitary transformation, i.e., the state $|\psi>$ of the system at time t_1 is related to the state $|\psi>$ at time t_2 by a unitary operator U which depends only on the initial (t_1) and end (t_2) times:

$$|\psi(t_2)>= U(t_2, t_1)|\psi(t_1)> . \tag{15.24}$$

Note that, $<\psi(t_2)|\psi(t_2)>=<\psi(t_1)|U^+U|\psi(t_1)>=<\psi(t_1)|\psi(t_1)>= 1$, i.e., the state vector stays normalized at all times.

In non-relativistic quantum-mechanics, postulate 2 is sometimes formulated as a prelude to the Schrödinger equation which describes the time evolution of the state of a closed system characterized by a Hamiltonian H

$$i\hbar\frac{d}{dt}|\psi>= H|\psi> . \tag{15.25}$$

If the Hamiltonian does not depend explicitly on time, a formal solution of the Schrödinger equation can be written as follows

$$|\psi(t_2)>= e^{\frac{-i}{\hbar}H(t_2-t_1)}|\psi(t_1)>, \tag{15.26}$$

and since H is Hermitian, the operator U defined as

$$U(t_2, t_1) = e^{\frac{-i}{\hbar}H(t_2-t_1)} \tag{15.27}$$

is unitary.

Exercises:

- Prove that $U(t_2, t_1) = e^{\frac{-i}{\hbar}H(t_2-t_1)}$ is unitary if H is Hermitian.
- Calculate the explicit form of the unitary operator for the case of a spin-1/2 particle (electron) in a spatially uniform and time independent magnetic field along the z-axis.

Solution

The Hamiltonian for this problem is given by

$$H = -\vec{\mu}.\vec{B}, \tag{15.28}$$

where $\vec{\mu}$ is the magnetic moment of the electron, which, in a bulk semiconductor with gyromagnetic factor g, is given by

$$\vec{\mu} = -\frac{g}{2}(\frac{e}{m_0})\vec{S}, \tag{15.29}$$

where

$$\vec{S} = \frac{\hbar}{2}\vec{\sigma}. \tag{15.30}$$

With $\vec{B}$ along the z-axis, the Hamiltonian becomes

$$H = -\vec{\mu}.\vec{B} = \frac{g\mu_B B}{2}\sigma_z = \frac{g}{2}\mu_B B \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \tag{15.31}$$

where $\mu_B = \frac{e\hbar}{2m_0}$ is the Bohr magneton. So

$$U(t_2 - t_1) = e^{-\frac{i}{\hbar}H(t_2-t_1)} = e^{-\frac{i}{\hbar}\frac{g\mu_B B}{2}\sigma_z(t_2-t_1)}. \tag{15.32}$$

Using the operator identity (3.52) proven in Chapter 3, we get

$$exp(\theta\sigma_z) = \begin{pmatrix} e^{\theta} & 0 \\ 0 & e^{-\theta} \end{pmatrix}, \tag{15.33}$$

where

$$\theta = -\frac{i}{\hbar}\frac{g\mu_B B}{2}\triangle t, \tag{15.34}$$

with $\triangle t = t_2 - t_1$.

Therefore,

$$U(t_2, t_1) = U(t_2 - t_1) = U(\triangle t) = \begin{pmatrix} e^{-\frac{i}{\hbar}\frac{g\mu_B B}{2}\triangle t} & 0 \\ 0 & e^{\frac{i}{\hbar}\frac{g\mu_B B}{2}\triangle t} \end{pmatrix}. \tag{15.35}$$

U is unitary, as easily checked.

Postulate 3: Quantum Measurements

The quantum projective approach of Von Neumann describes the wavefunction collapse associated with measurements performed on quantum-mechanical systems. According to this approach, a projective measurement is described by an observable M, which is a Hermitian operator on the state space of the system. Measurements obey the following set of rules:

- Since the operator M is Hermitian, it has a spectral decomposition

$$M = \sum_m m P_m, \tag{15.36}$$

  where P_m is the projector onto the eigenspace of M with eigenvalue m and corresponding eigenvector $|m>$. P_m is explicitly given by the outer product

$$P_m = |m><m|. \tag{15.37}$$

- The possible outcomes of the measurement of the observable M correspond to the eigenvalues m of the observable.

- Upon measuring M on a quantum-mechanical system characterized by the ket $|\psi>$, the probability of getting result m is given by

$$p(m) = <\psi|P_m|\psi> = <\psi|m><m|\psi> = |<\psi|m>|^2. \tag{15.38}$$

- After the measurement is made, if outcome m has occurred, the state of the quantum system immediately after the measurement is given by (or has collapsed to)

$$P_m|\psi> / \sqrt{p(m)} = \frac{<m|\psi>|m>}{\sqrt{|<\psi|m>|^2}}, \tag{15.39}$$

  where $\sqrt{p(m)}$ is introduced to renormalize the state after collapse occurs.

A few problems on quantum projective measurements involving the spin components of a single qubit are discussed in Chapters 3 and 4.

Exercise: A useful property of projection measurements

In order to calculate the average value of repeated measurements of the observable M on a quantum-mechanical system in the state $|\psi>$, we must perform the summation

$$E(M) = \sum_m m p(m), \tag{15.40}$$

which is what we expect from the definition of average value associated with a random variable using probability theory. Plugging in the definition of $p(m) =< \psi|P_m|\psi >$

$$E(M) = \sum_m m < \psi|P_m|\psi > =< \psi| \sum_m mP_m|\psi > =< \psi|M|\psi > . \quad (15.41)$$

The latter expression can usually be calculated easily for a given operator M and state $|\psi >$.

Similarly, the standard deviation of a large number of measurements of the observable M is defined by

$$\Delta(M) = [< \psi|(M - < M >)^2|\psi >]^{\frac{1}{2}}, \quad (15.42)$$

where

$$< M > \doteq < \psi|M|\psi > . \quad (15.43)$$

Therefore,

$$\Delta(M) = \sqrt{< \psi|(M^2 - 2 < M > M + < M >^2)|\psi >} = \sqrt{< M^2 > - < M >^2}, \quad (15.44)$$

where $< M^2 > =< \psi|(M^2|\psi >$.

The standard deviation $\Delta(M)$ can therefore also be calculated easily once the operator M and the ket $|\psi >$ are known.

The Heisenberg Uncertainty Principle

For any two Hermitian operators A, B associated with physical observables, the average value of the product AB in a quantum state $|\psi >$ will be some complex number $< \psi|AB|\psi > = x + iy$ where x, y are real. Since AB is Hermitian, $< \psi|BA|\psi > = x - iy$ and the following equalities hold

$$< \psi|[A, B]|\psi > =< \psi|AB|\psi > - < \psi|BA|\psi > = (x + iy) - (x - iy) = 2iy, \quad (15.45)$$

and

$$< \psi|\{A, B\}|\psi > =< \psi|AB|\psi > + < \psi|BA|\psi > = (x + iy) + (x - iy) = 2x. \quad (15.46)$$

Hence,

$$| < \psi|[A, B]|\psi > |^2 + | < \psi|\{A, B\}|\psi > |^2 = 4| < \psi|AB|\psi > |^2 = 4(x^2 + y^2). \quad (15.47)$$

Using the Cauchy-Schwartz inequality

$$| < v|w > |^2 \leq < v|v >< w|w >, \quad (15.48)$$

with $|v > = A|\psi >$ and $|w > = B|\psi >$, we get

$$| < \psi|AB|\psi > |^2 \leq < \psi|A^2|\psi >< \psi|B^2|\psi >, \quad (15.49)$$

which leads to

$$| < \psi|[A, B]|\psi > |^2 \leq 4 < \psi|A^2|\psi >< \psi|B^2|\psi > . \tag{15.50}$$

Using the definition of the standard deviation associated with a measurement of an observable M and defining two new operators C and D such that

$$C = A - < A >, D = B - < B >, \tag{15.51}$$

we get

$$< C >= 0, \Delta(C) = \sqrt{< C^2 >} = \Delta(A), \tag{15.52}$$

and

$$< D >= 0, \Delta(D) = \sqrt{< B^2 >} = \Delta(B). \tag{15.53}$$

Furthermore, $[A, B] = [C, D]$ and therefore the inequality (15.50) can be rewritten

$$\Delta(C)\triangle(D) \geq \frac{| < \psi|[C, D]|\psi > |}{2}. \tag{15.54}$$

This is the generalized form of the Heisenberg uncertainty principle.

Exercises:

In the derivation above, it was assumed that, for any two Hermitian operators A and B, the average value of the product AB in a quantum state $|\psi >$ is equal to some complex number $< \psi|AB|\psi >= x + iy$ where x, y are real. Find the values of x and y.

- For the case where $A = \sigma_x$ and $B = \sigma_y$, and the ket $|\psi >$ equal to $|0 >$, $|1 >$ and $\frac{1}{\sqrt{2}}(|0 > + |1 >)$.
- Repeat the previous exercise if $A = \mathbf{H}$, the Hadamard matrix (see Eq.(3.41)) and B $= \sigma_y$.
- Find the value of $x + iy$ for an electron in the ground state of the one-dimensional harmonic oscillator in a box taking $A = x$ (position) and $B = p_x$ (momentum). The wavefunction in the ground state of a one-dimensional harmonic oscillator described by a potential energy $V(x) = \frac{m}{2}\omega^2 x^2$ is given by

$$\phi_0(x) = (\frac{m\omega}{\pi\hbar})^{1/4} e^{-\frac{1}{2}\frac{m\omega}{\hbar}x^2}. \tag{15.55}$$

- Repeat the previous problem if the electron is in the first excited state of the one-dimensional harmonic oscillator. The wavefunction in the first excited of the one-dimensional harmonic oscillator is given by

$$\phi_1(x) = [\frac{4}{\pi}(\frac{m\omega}{\hbar})^3]^{1/4} x e^{-\frac{1}{2}\frac{m\omega}{\hbar}x^2}. \tag{15.56}$$

15.4.1 Interpretation of the Heisenberg Uncertainty Principle

From a practical standpoint, the meaning of the inequality (15.54) is as follows:

- We must first prepare a physical system a large number of times (or ensemble) in state $|\psi>$,
- Then, perform measurements of the observable C on a fraction of the ensemble and of the observable D on the rest of the ensemble.
- When a meaningful number of measurements have been performed so that a standard deviation of these measurements can be calculated, the experimental results for the standard deviation $\Delta(C)$ of the measurements of the observable C and the standard deviation $\Delta(D)$ of the measurements of the observable D will satisfy the inequality

$$\Delta(C)\Delta(D) \geq \frac{|<\psi|[C,D]|\psi>|}{2}. \tag{15.57}$$

Example 1: Applying the inequality in the case of the position and momentum operators x and p_x, and using the fact that $[x, p_x] = i\hbar$, we get

$$\Delta(x)\triangle(p_x) \geq \frac{\hbar}{2}, \tag{15.58}$$

which is one of the early forms of the Uncertainty Principle published by Heisenberg.

Example 2:
Consider the Pauli matrices σ_x and σ_y and show that the inequality (15.54) leads to

$$\Delta(S_x)\Delta(S_y) \geq \frac{\hbar^2}{4}. \tag{15.59}$$

This can be easily shown starting with the inequality (15.54) and the following property of the Pauli matrices proved in Chapter 2:

$$[\sigma_x, \sigma_y] = 2i\sigma_z. \tag{15.60}$$

So, $\Delta(S_x), \Delta(S_y)$ must both be strictly greater than 0. This can be verified by direct calculation, as shown in the next concrete example.

Calculate both sides of the inequality (15.54) for an ensemble of qubits prepared in the quantum state $|0>$, eigenvector of the Pauli matrix σ_z with eigenvalue $+1$, i.e.,

$$|0> = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \tag{15.61}$$

for which,

$$< \sigma_x >= (1\ 0) \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = 0 \tag{15.62}$$

and

$$< \sigma_y >= (1\ 0) \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = 0. \tag{15.63}$$

Furthermore, since $< {\sigma_x}^2 > = < {\sigma_y}^2 > = \frac{\hbar^2}{4}$, we get

$$\Delta S_x \Delta S_y = \frac{\hbar^2}{4}, \tag{15.64}$$

and the inequality (15.54) is an equality in this case.

Example 3: Heisenberg Uncertainty Principle for the 1D particle in a box

Starting with the eigenstates and corresponding eigenvalues for the simple problem of a particle in a one-dimensional quantum box of size W,

- (a) Calculate the average values $< x >$ and $< p_x >$ for an electron prepared in each of the eigenstates of the particle in a box.
- (b) Calculate the standard deviations Δx, Δp_x for an electron prepared in each of the eigenstates of the particle in a box.
- (c) Show that for large values of the quantum number n characterizing the eigenstates, the standard deviation Δx reduces to its classical value.
- (d) Calculate $\sqrt{\Delta x \Delta p_x}$ for each of the eigenstates found in part (a) and show that Heisenberg Uncertainty relation is satisfied.

Solution

The solution to the Schrödinger equation for a particle in a one-dimensional box is one of the simplest problems in quantum mechanics and is given by

$$\phi_n(x) = \sqrt{\frac{2}{W}} sin(\frac{n\pi x}{W}), \tag{15.65}$$

where n is an integer (n = 1,2,3,...) and W is the width of the box. The corresponding eigenvalues are given by

$$E_n = n^2 \frac{\hbar^2 \pi^2}{2mW^2}. \tag{15.66}$$

(a) Using these results, the following average values are found

- $$< x >= \int_0^W x|\phi_n(x)|^2 dx = \frac{W}{2}. \tag{15.67}$$

  This is to be expected since all eigenstates produce symmetrical probability distributions $|\phi_n|^2$ with respect to the center of the box.

- $$< p_x >= \int_0^W dx\phi_n(x)\frac{\hbar}{i}\frac{d}{dx}\phi_n(x) = 0. \tag{15.68}$$

  This is expected since $\phi_n(x)$ and $\frac{d}{dx}\phi_n(x)$ are sine and cosine functions, respectively, and are therefore orthogonal to each other. Classically, this corresponds to the fact that a particle traveling back and forth between the two walls of the box will have the same magnitude of its momentum but opposite signs when traveling from left to right and right to left. Therefore, the average momentum will be zero when trapped inside a box.

(b) Standard deviations

- $$< \Delta x^2 >= \int_0^W dx x^2 \phi_n^2(x) - \left[\frac{W}{2}\right]^2. \tag{15.69}$$

  So
  $$(\Delta x)^2 = \frac{2}{W}\int_0^W dx x^2 sin^2\left(\frac{n\pi x}{W}\right) - \frac{W^2}{4}. \tag{15.70}$$

  Therefore
  $$(\Delta x)^2 = \frac{W^2}{12}\left[1 - \frac{6}{(n\pi)^2}\right]. \tag{15.71}$$

- $$(\Delta p_x)^2 =< p_x^2 > - < p_x >^2 =< p_x^2 > . \tag{15.72}$$

Hence
$$(\Delta p_x)^2 = \int_0^W dx\phi_n(x)p_x^2\phi_n(x)dx. \tag{15.73}$$

But $p_x^2\phi_n = 2mE_n\phi_n(x)$, so

$$(\Delta p_x)^2 = 2mE_n\int_0^W dx|\phi_n|^2 = 2m\left(\frac{\hbar^2 n^2\pi^2}{2mW^2}\right) = \frac{\hbar^2 n^2\pi^2}{W^2}. \tag{15.74}$$

(c) As n increases, we have from part (b)

$$(\Delta x)^2 = \frac{W^2}{12}. \tag{15.75}$$

We compare this to its classical limit which is equal to

$$\Delta x^2 = \frac{1}{W}\int_0^W (x - \frac{W}{2})^2 dx = \frac{W^2}{12}. \tag{15.76}$$

So indeed the quantum-mechanical result converges to its classical value for eigenstates with large quantum numbers.

(d) Value of $\sqrt{\Delta x \Delta p_x}$:

Using the previous results, we get

$$\sqrt{\Delta x \Delta p_x} = \frac{\hbar}{2}\frac{n\pi}{\sqrt{3}}[1 - \frac{6}{n^2\pi^2}]^{\frac{1}{2}}, \tag{15.77}$$

which can be shown to be greater than $\frac{\hbar}{2}$ for all integers ≥ 1. The smallest value of the right hand side in the above equation is for $n = 1$, in which case

$$\sqrt{\Delta x \Delta p_x} = 0.568\hbar, \tag{15.78}$$

which is slightly larger than $\frac{\hbar}{2}$.

15.4.2 Time evolution of expectation values: the Ehrenfest Theorem

Ehrenfest was the first to derive an equation to describe the time evolution of the expectation value of a time-dependent observable for a quantum-mechanical system described by a ket $|\phi(t)>$. Using Dirac's notations,

$$\frac{d}{dt} < \phi(t)|A(t)|\phi(t) >= [\frac{d}{dt} < \phi(t)|]A(t)|\phi(t) > + < \phi(t)|A(t)[\frac{d}{dt}|\phi(t) >] + < \phi(t)|\frac{dA(t)}{dt}|\phi(t) > . \tag{15.79}$$

Taking into account that $|\phi(t) >$ satisfies the Schroedinger equation, we get

$$\frac{d}{dt}|\phi(t) >= \frac{1}{i\hbar}H(t)|\phi(t) > . \tag{15.80}$$

The conjugate transpose of this last equation is given by

$$\frac{d}{dt} < \phi(t)| = -\frac{1}{i\hbar}H(t) < \phi(t)|. \tag{15.81}$$

Using the last two equations, Equation (15.79) can be rewritten as follows

$$\frac{d}{dt} < \phi(t)|A(t)|\phi(t) >= \frac{1}{i\hbar}\frac{d}{dt} < \phi(t)|[A(t)H(t) - H(t)A(t)]|\phi(t) > + < \phi(t)|\frac{dA(t)}{dt}|\phi(t) > \tag{15.82}$$

or

$$\frac{d}{dt} < A >= \frac{1}{i\hbar} < [A, H(t)] > + < \frac{dA}{dt} > . \tag{15.83}$$

This last equation is referred to as *Ehrenfest's theorem.*

Example 1:

We first illustrate the application of Ehrenfest's theorem to a spinless particle whose Hamiltonian is given by

$$H = \frac{p^2}{2m} + V(r), \tag{15.84}$$

where $V(r)$ is the potential energy of the particle.

Applications of Ehrenfest's theorem with operator A equal to the three components of the position operator lead to

$$\frac{d}{dt} < r >= \frac{1}{i\hbar} < [r, H] >= \frac{1}{i\hbar} < [r, \frac{p^2}{2m}] > . \tag{15.85}$$

Similarly, if A is equal to the three components of the momentum operator, we get

$$\frac{d}{dt} < p >= \frac{1}{i\hbar} < [p, H] >= \frac{1}{i\hbar} < [p, V(r)] > . \tag{15.86}$$

Using appropriate identities to calculate commutators of function of operators [17, 18]

$$[r, \frac{p^2}{2m}] = \frac{i\hbar}{m} p, \tag{15.87}$$

and

$$[p, V(r)] = -i\hbar \nabla V(r) \tag{15.88}$$

and combining that with the Ehrenfest theorem ultimately lead to

$$\frac{d}{dt} < \vec{r} >= \frac{1}{m} < p >, \tag{15.89}$$

and

$$\frac{d}{dt} < \vec{p} >= - < \nabla V(r) > . \tag{15.90}$$

The last two equations are similar to the classical equations of motion (Newton's equations) describing the time evolution of the momentum of a particle except that the latter is replaced by the average value of the corresponding quantum-mechanical operator over the state $|\phi(t) >$ of the system. The force acting on the particle is calculated as the average over the state of the particle of the gradient of the potential energy.

Example 2:

In Chapter 4, we invoked the Ehrenfest's theorem to describe the time evolution of a spin-1/2 particle in a spatially uniform time dependent magnetic field whose Hamiltonian is given by -$(g/2)\mu_B\vec{\sigma}\cdot\vec{B}$. Here, we provide the proof.

According to Ehrenfest's theorem, the time evolution of the averages of the spin components are given by

$$\frac{d}{dt} < \sigma_x >= \frac{1}{i\hbar}\frac{g}{2}\mu_B(B_y < [\sigma_x, \sigma_y] > + B_z < [\sigma_x, \sigma_z] >), \tag{15.91}$$

$$\frac{d}{dt} < \sigma_y >= \frac{1}{i\hbar}\frac{g}{2}\mu_B(B_x < [\sigma_y, \sigma_x] > + B_z < [\sigma_y, \sigma_z] >), \tag{15.92}$$

and

$$\frac{d}{dt} < \sigma_z >= -\frac{1}{i\hbar}\frac{g}{2}\mu_B(B_x < [\sigma_z, \sigma_x] > + B_y < [\sigma_z, \sigma_y] >). \tag{15.93}$$

Using the relations in Problem 2.4, we get

$$\frac{d}{dt} < \sigma_x >= \frac{g\mu_B}{\hbar}(B_y < \sigma_z > - B_z < \sigma_y >), \tag{15.94}$$

$$\frac{d}{dt} < \sigma_y >= \frac{g\mu_B}{\hbar}(B_z < \sigma_x > - B_x < \sigma_z >), \tag{15.95}$$

and

$$\frac{d}{dt} < \sigma_z >= \frac{g\mu_B}{\hbar}(B_x < \sigma_y > - B_y < \sigma_x >). \tag{15.96}$$

The reader can easily show that the last three equations can be written as

$$\frac{d < \vec{\sigma} >}{dt} = \frac{g\mu_B\vec{B}}{\hbar} \times < \vec{\sigma} >, \tag{15.97}$$

which leads to the familiar Larmor precession equation

$$\frac{d\vec{S}}{dt} = \frac{g\mu_B\vec{B}}{\hbar} \times \vec{S}, \tag{15.98}$$

since $\vec{S}_{op} = (\hbar/2)\vec{\sigma}$.

15.5 Some elements of semiconductor physics: particular applications in nanostructures

15.5.1 Density of states: bulk (3-D) to quantum dot (0-D)

Consider the quantum confined geometries shown in Fig. 15.3 (2D: two-dimensional electron gas, 1D: one-dimensional electron gas, 0D: three-dimensional

quantum box). Calculate the energy dependence of the *density of states* in these structures and compare them to that of the 3D bulk sample shown in the upper left corner in Fig. 15.3.

Solution

3D: Consider a uniform homogeneous bulk piece of semiconductor whose conduction band has a parabolic $E - \vec{k}$ relationship with the bottom at E_{co}, as shown in Figure 15.4:

$$E(\vec{k}) = E_{co} + \frac{\hbar^2 k^2}{2m^*}. \tag{15.99}$$

The solutions of the 3D effective mass Schrödinger equation are of the form of plane waves

$$\phi_k(\vec{r}) = \frac{1}{\sqrt{\Omega}} e^{i\vec{k}\vec{r}}, \tag{15.100}$$

normalized over a volume $\Omega = L^3$, where L is the length of the side of a cube large compared to the lattice unit cell. Assuming periodic boundary conditions for $\phi_k(\vec{r})$, i.e.,

$$\phi_k(x + L, y + L, z + L) = \phi_k(x, y, z), \tag{15.101}$$

the allowed values of $\vec{k} = (k_x, k_y, k_z)$ are given by

$$k_x = n_x \frac{\pi}{L}, \tag{15.102}$$

$$k_y = n_y \frac{\pi}{L}, \tag{15.103}$$

and

$$k_z = n_z \frac{\pi}{L}, \tag{15.104}$$

where n_x, n_y, n_z are integers.

The density of electrons at location $\vec{r}$ can then be calculated as follows

$$\rho(\vec{r}) = \frac{N}{\Omega} = \sum_{\vec{k}} f(E_k) |\phi_k(\vec{r})|^2, \tag{15.105}$$

where $f(E_k)$ is the Fermi-Dirac distribution function. We can assume that the carrier statistics is governed by the Fermi-Dirac distribution as long as the system is in *equilibrium* (e.g., no current flows and no light is shining on it generating electron-hole pairs).

Each electron eigenstate occupies a volume $\frac{(2\pi)^3}{\Omega}$ in $\vec{k}$-space. Therefore, in a volume of size $d^3\vec{k}$, we have a number of electron eigenstates equal to

$$2\frac{\Omega}{(2\pi)^3} d^3\vec{k}, \tag{15.106}$$

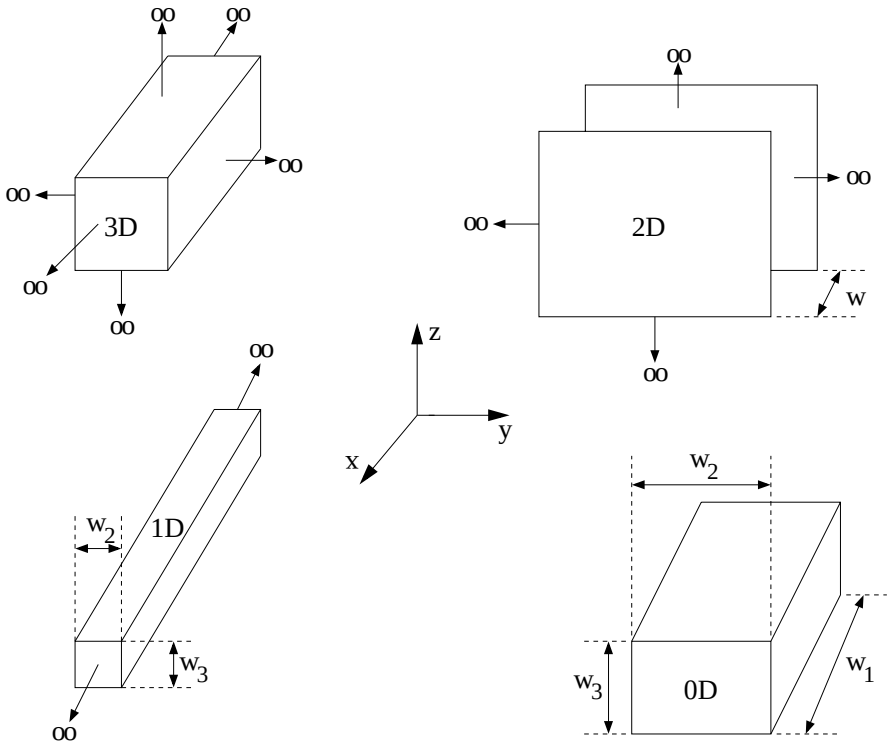


FIGURE 15.3
Illustration of the formation of a quantum dot (bottom right figure) through the gradual squeezing of a bulk piece of semiconductor (upper left). When the dimension of the bulk structure is reduced in one direction to a size comparable to the de Broglie wavelength, the resulting electron gas is referred to as a two-dimensional electron gas (2 DEG) because the carriers are free to move in the y and z directions only. If quantum confinement occurs in two directions, as illustrated in the bottom left figure, the resulting electron gas is referred to as a one-dimensional electron gas (1 DEG) since an electron in this structure is free to move in the x direction only. If confinement is imposed in all three directions, we get a quantum dot (0 DEG).

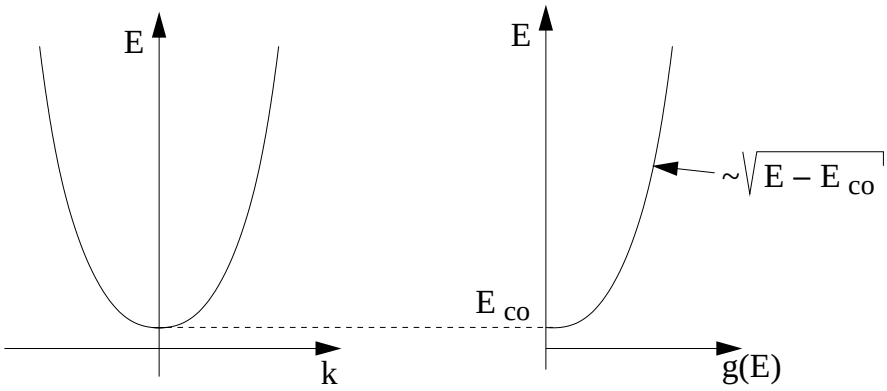


FIGURE 15.4
(Left) Parabolic energy dispersion relation close to the bottom of the conduction band (E_{c0}) of a typical semiconductor. (Right) Corresponding energy dependence of the three-dimensional density of states in a bulk semiconductor.

where the extra factor 2 has been added to take into account the spin degeneracy of each eigenstate in $\vec{k}$-space, as required by the Pauli Exclusion principle. For a large value of Ω, the $\sum_{\vec{k}}$ in Equation (15.105) can be replaced by an integral and we obtain

$$\rho(\vec{r}) = \frac{1}{4\pi^3} \int d^3\vec{k} f(E_k), \tag{15.107}$$

which is spatially invariant.

Since $f(E_k)$ is spherically symmetric in $\vec{k}$-space, the last integration can be easily performed using spherical coordinates leading to

$$\rho = \int_{E_{co}}^{+\infty} dE g_{3D}(E) f(E), \tag{15.108}$$

where $g_{3D}(E)$ is by definition the three-dimensional density of states and is given by

$$g_{3D}(E) = \frac{k^2}{\pi^2(\frac{dE}{dk})} = \frac{m^* k}{\pi^2 \hbar^2}, \tag{15.109}$$

where we have used the dispersion relation in Equation (15.99) to arrive at the last equality.

Using the E-k relationship (15.99) once again to express k in terms of E, we get the well-known result for the 3-D density of states:

$$g_{3D}(E) = \frac{m^*}{\pi^2 \hbar^3} \sqrt{2m^*(E - E_o)}, \tag{15.110}$$

whose energy dependence is illustrated in Figure 15.4.

Using the above equation in Equation (15.107), the electron density in a bulk sample is given by

$$\rho = \frac{2}{\pi} N_c F_{\frac{1}{2}}(\xi), \tag{15.111}$$

where

$$N_c = \frac{1}{4\hbar^3}(\frac{2m^* k_B T}{\pi})^{\frac{3}{2}}, \tag{15.112}$$

and

$$\xi = \frac{(E_F - E_{co})}{k_B T}. \tag{15.113}$$

In equation (15.111), $F_{\frac{1}{2}}$ is the Fermi-Dirac integral of index $\frac{1}{2}$:

$$F_{\frac{1}{2}}(\xi) = \int_{E_{co}}^{+\infty} \frac{dE\sqrt{E - E_{co}}}{[1 + e^{\frac{E-E_F}{kT}}]}. \tag{15.114}$$

2D: Next, we generalize the derivation above to determine the two-dimensional density of states in a 2 DEG. In this case, the electron density is calculated as follows

$$\rho(\vec{r}) = \sum_m \sum_{k_y,k_z} f_o(E_m)|\phi_{m,k_y,k_z}(\vec{r})|^2. \tag{15.115}$$

The eigenfunctions and corresponding eigenvalues of the Schrödinger equation are given by:

$$\phi_{m,k_y,k_z}(\vec{r}) = \frac{1}{\sqrt{A}} e^{ik_y y} e^{ik_z z} \xi_m(x), \tag{15.116}$$

where

$$E_{m,k_y,k_z} = E_m + \frac{\hbar}{2m^*}(k_y^2 + k_z^2), \tag{15.117}$$

and

$$A = L_y L_z \tag{15.118}$$

is a normalization area to describe the in-plane free motion of carriers in the (y, z) directions, x being the direction of quantum confinement of the well. The wavefunctions $\xi_m(x)$ are solutions of the one-dimensional Schrödinger equation and depend on the potential confinement $E_c(x)$ in the x-direction

$$-\frac{\hbar^2}{2m^*}\frac{d^2\xi_m(x)}{dx^2} + E_c(x)\xi_m(x) = E_m\xi_m(x). \tag{15.119}$$

Each $\xi_m(x)$ is assumed to be normalized and has a corresponding eigenvalue E_m.

Therefore,

$$\rho(\vec{r}) = \sum_m \sigma_m |\xi_m(x)|^2, \tag{15.120}$$

where

$$\sigma_m = \sum_{k_y,k_z} \frac{1}{A} f_o(E_{m,k_y,k_z}). \tag{15.121}$$

Converting the $\sum_{k_y,k_z}$ to an integral following the 3-D case, we get

$$\sum_{k_y,k_z} = 2\left(\frac{A}{4\pi^2}\right)\int d^2\vec{k}. \tag{15.122}$$

Using polar coordinates in the (k_y, k_z) plane

$$\sigma_m = \int_0^{2\pi} \frac{d\phi}{2\pi^2} \int_0^{+\infty} dk k f_o(E_{m,k_y,k_z}), \tag{15.123}$$

and since $f_o(E_{m,k_y,k_z})$ is independent of ϕ,

$$\sigma_m = \int_0^{+\infty} \frac{kdk}{\pi} f_o(E_{m,k_y,k_z}). \tag{15.124}$$

Using the dispersion relationship of the subbands in the well, we get

$$dE_{m,k_y,k_z} = \frac{\hbar}{n^*} k dk, \tag{15.125}$$

and σ_m becomes

$$\sigma_m = \int_{E_m}^{+\infty} dE g_{2D}(E) f_o(E), \tag{15.126}$$

where

$$g_{2D}(E) = \frac{m^*}{\pi\hbar^2} \tag{15.127}$$

is *independent of energy* and is the density of states in each subband in the well.

Substituting the expression for the Fermi-Dirac factor $f_o(E)$, σ_m can be calculated exactly,

$$\sigma_m = \frac{m^*}{\pi\hbar^2} k_B T ln(1 + e^{\frac{E_m - E_F}{k_B T}}). \tag{15.128}$$

This analytical expression for σ_m is valid for any shape of the confining potential in the x direction. This quantity determines the sheet electron concentration in a 2-DEG.

1D:

If we have confinement in the y-z plane and free motion of carriers is allowed in the x-direction, then

$$\rho(\vec{r}) = \sum_{k_x} \sum_{n} \sum_{m} f_o(E_{n,m}) |\phi_{k_x,n,m}(\vec{r})|^2 \tag{15.129}$$

where

$$\phi_{k_x,n,m}(\vec{r}) = \frac{1}{\sqrt{L}} e^{ik_x x} \xi_{n,m}(y,z), \tag{15.130}$$

where L is a normalization factor of the plane wave moving along the x-direction and $\xi_{n,m}(y,z)$ are the solutions of the two-dimensional Schrödinger equation

$$-\frac{\hbar^2}{2m^*}\left(\frac{d^2}{dy^2}+\frac{d^2}{dz^2}\right)\xi_{n,m}(y,z)+E_c(y,z)\xi_{n,m}(y,z) = E_{n,m}\xi_{n,m}(x,y). \tag{15.131}$$

Here, n, m are quantum numbers characterizing the quantization in the y and z directions. They are also called *transverse subband indices.*

The energy dispersion relationship in each subband characterized by the two quantum numbers (n, m) is given by

$$E_{k_x,n,m} = E_{n,m} + \frac{\hbar^2}{2m^*}k_x^2. \tag{15.132}$$

Therefore, in this 1-DEG, the electron density is invariant in the x-direction

$$\rho(y,z) = \sum_{n,m} \sigma_{n,m} |\xi_{n,m}(y,z)|^2, \tag{15.133}$$

where

$$\sigma_{n,m} = \sum_{k_x} \frac{1}{L} f_o(E_{k_x,n,m}). \tag{15.134}$$

Converting the sum over k_x into an integral, i.e.,

$$\sum_{k_x} = 2\left(\frac{L}{2\pi}\right)\int dk_x, \tag{15.135}$$

we get

$$\sigma_{n,m} = \frac{1}{\pi}\int_{-\infty}^{+\infty} dk_x f_o(E_{k_x,n,m}) = \frac{2}{\pi}\int_0^{+\infty} dk_x f_o(E_{k_x,n,m}). \tag{15.136}$$

$$\tag{15.137}$$

Using the dispersion relation ($E - k_x$ relation), we get

$$dE_{k_x,n,m} = \frac{\hbar}{m^*}k_x dk_x; \tag{15.138}$$

hence

$$\sigma_{n,m} = \int_{E_{n,m}}^{+\infty} dE g_{1D}(E) f_o(E), \tag{15.139}$$

where

$$g_{1D}(E) = \frac{1}{\pi}\sqrt{\frac{2m^*}{\hbar^2}}\frac{1}{\sqrt{E - E_{n,m}}}, \tag{15.140}$$

which is the expression for the one-dimensional density of states in each subband in the quantum wire. It diverges at $E = E_{n,m}$, the threshold energy for free propagation in that subband.

0D: In that case, we are dealing with a quantum box with quantum confinement in all three directions.

$$\rho(\vec{r}) = \sum_{n,m,l} f_o(E_{n,m,l})|\phi_{n,m,l}(\vec{r})|^2, \tag{15.141}$$

where $\phi_{n,m,l}$ are the solutions of the three-dimensional Schrödinger equation for the $E_c(x, y, z)$ representing the quantum confinement in all three directions. The indices (n, m, l) are three quantum numbers characterizing the eigenstates of the Schrödinger equation.

We can write

$$\rho(\vec{r}) = \sum_{n,m,l} \sigma_{n,m,l}|\phi_{n,m,l}(\vec{r})|^2, \tag{15.142}$$

with

$$\sigma_{n,m,l} = \int_0^{+\infty} dE g_{0D}(E) f_o(E). \tag{15.143}$$

Therefore the 0-dimensional density of states is simply

$$g_{0D}(E) = 2\sum_{n,m,l} \delta(E - E_{n,m,l}) \tag{15.144}$$

where δ is the Dirac delta function and the factor 2 has been included since each $E_{n,m,l}$ state can be occupied by two electrons with opposite spin.

Example 1: Electron sheet concentration in a quantum well

(a) Show that the sheet carrier concentration in a HEMT device (which has a 2-DEG in the channel) is given by

$$n_s = \frac{m^*}{\pi\hbar^2}k_B T ln\left[(1 + e^{\frac{E_F - E_1}{k_B T}})(1 + e^{\frac{E_F - E_2}{k_B T}})\right], \tag{15.145}$$

when only two subbands are occupied. Here, E_1 and E_2 are the bottom energies of the first two subbands.

(b) Starting with the result of the part (a), show that at low temperature

$$n_s = \frac{m^*}{\pi\hbar^2}(E_F - E_1), \tag{15.146}$$

when the second subband is unoccupied and

$$n_s = \frac{m^*}{\pi\hbar^2}(E_2 - E_1) + 2\frac{m^*}{\pi\hbar^2}(E_F - E_2), \tag{15.147}$$

when both subbands are occupied.

Solution

The electron concentration in the 2 DEG formed at the heterointerface between the high and low bandgap materials in a HEMT structure (see Fig. 12.6) is given by

$$\rho(x) = \sum_m \sigma_m |\xi_m(x)|^2, \tag{15.148}$$

where

$$\sigma_m = \frac{m^*(k_B T)}{\pi\hbar^2} ln(1 + e^{-\frac{(E_m - E_F)}{k_B T}}). \tag{15.149}$$

The sheet carrier concentration in the 2 DEG is given by

$$n_s = \int_{-\infty}^{+\infty} \rho(x)dx. \tag{15.150}$$

If the wavefunctions $\xi_m(x)$ are normalized, i.e.,

$$\int_{-\infty}^{+\infty} |\xi_m(x)|^2 dx = 1, \tag{15.151}$$

the sheet carrier concentration is given by the simple formula

$$n_s = \sum_m \sigma_m. \tag{15.152}$$

If only one subband in the 2 DEG is occupied,

$$n_s = k_B T ln(1 + e^{\frac{E_F - E_1}{k_B T}}). \tag{15.153}$$

If $k_B T << E_F - E_o$,

$$n_s = \frac{m^*}{\pi\hbar^2}(E_F - E_1). \tag{15.154}$$

When the second subband is occupied (but the third one unoccupied),

$$n_s = \frac{m^*}{\pi\hbar^2} k_B T \left[ln(1 + e^{\frac{E_F - E_1}{k_B T}}) + ln(1 + e^{\frac{E_F - E_1}{k_B T}}) \right], \tag{15.155}$$

which can also be written as follows

$$n_s = \frac{m^*}{\pi\hbar^2} k_B T ln \left[(1 + e^{\frac{E_F - E_1}{k_B T}})(1 + e^{\frac{E_F - E_2}{k_B T}}) \right]. \tag{15.156}$$

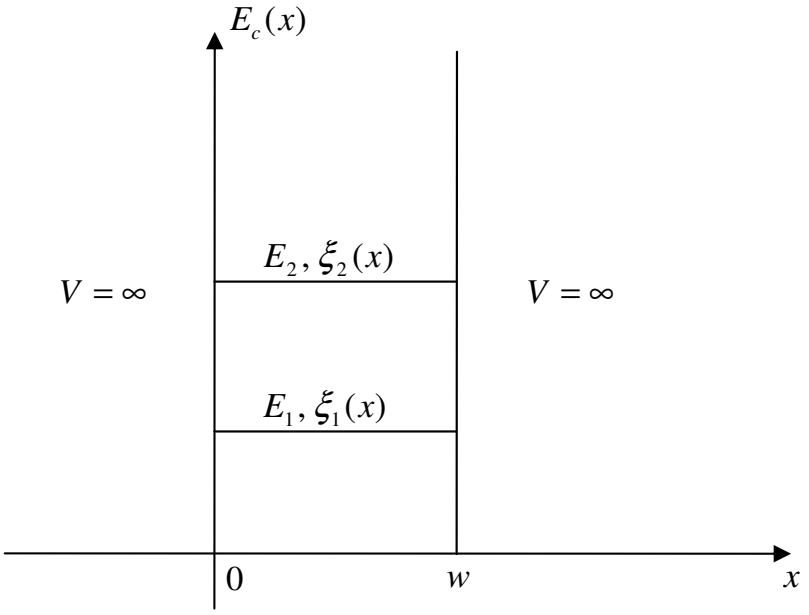


FIGURE 15.5
Confined states in a quantum well (2 DEG) of width w.

If $k_BT << E_F - E_1$, $E_F - E_2$, (i.e., at low temperature), then

$$ln\left(1 + e^{\frac{E_F - E_1}{k_BT}}\right) = \frac{E_F - E_1}{k_BT}, \tag{15.157}$$

and

$$ln\left(1 + e^{\frac{E_F - E_2}{k_BT}}\right) = \frac{E_F - E_2}{k_BT}. \tag{15.158}$$

Hence

$$n_s = \frac{m^*}{\pi\hbar^2}(E_F - E_o) + \frac{m^*}{\pi\hbar^2}(E_F - E_1) = \frac{m^*}{\pi\hbar^2}(2E_F - E_1 - E_2). \tag{15.159}$$

Example 2: Fermi level location in a Quantum Well

Consider a $100\mathring{A}$ wide potential well (quantum well or 2-DEG) with infinite walls at $T = 0K$. Assume all impurities are ionized (i.e., neglect carrier freeze out). Assume $m^* = 0.067m_0$ and calculate the location of the Fermi level for

- $N_D = 10^{17}cm^{-3}$,
- $N_D = 10^{19}cm^{-3}$.

Solution

Assuming that all impurities are ionized, the sheet carrier concentration in the well is given by

$$n_s = N_D W. \tag{15.160}$$

Therefore, for $N_D = 10^{17}$ and $10^{19} cm^{-3}$, n_s is equal to 10^{11} and $10^{13} cm^{-2}$, respectively.

At zero temperature, if the Fermi level E_F is between the N^{th} and $(N+1)^{th}$ subbands in the well, then

$$n_s = \frac{m^*}{\pi\hbar^2} \sum_{i=1}^{N} (E_F - E_i). \tag{15.161}$$

Hence

$$n_s = \frac{m^*}{\pi\hbar^2} \left(N E_F - \sum_{i=1}^{N} E_i \right), \tag{15.162}$$

which is a generalization of the results found in the previous example. Solving for E_F we get

$$E_F = \frac{1}{N} \left[\frac{N_D W}{(\frac{m^*}{\pi\hbar^2})} + \sum_{i=1}^{N} E_i \right]. \tag{15.163}$$

For a well surrounded by a infinite wall (particle in a box problem), the different eigenstates energies are given by

$$E_i = \frac{\hbar^2}{2m^*} (\frac{i\pi}{W})^2, \tag{15.164}$$

where i is an integer.

For $m^* = 0.067 m_o$ and $W = 100 \mathring{A}$, we find

$$E_i \simeq i^2 56 meV. \tag{15.165}$$

Therefore, the subband energy bottoms due to the particle-in-a-box confinement in two dimensions are given by the above equation.

For $N_D = 10^{17} cm^{-3}$, if we assume E_F is between E_1 and E_2, and $N = 1$ in Equation (15.163) above, we get

$$E_F = 59.78 meV, \tag{15.166}$$

which tells us that only one subband is occupied, as assumed.

For $N_D = 10^{19} cm^{-3}$, assuming E_F is between E_2 and E_3, the Fermi level is found to be

$$E_F = 329.4 meV, \tag{15.167}$$

which is below E_3. In this case, only two subbands are occupied. The number of subbands that are occupied is usually found by trial and error.

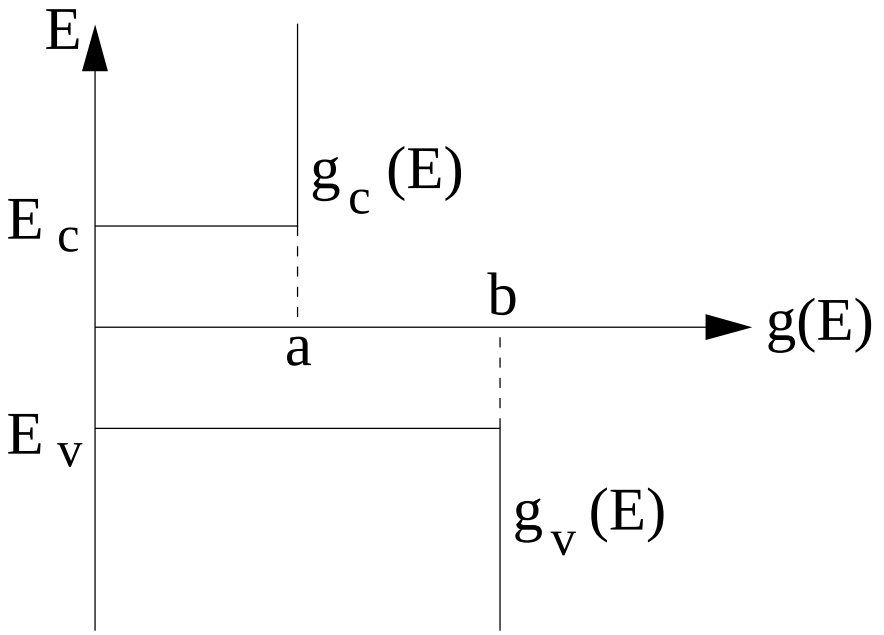


FIGURE 15.6
Density of states of electrons or holes in a two dimensional electron or hole gas (2-DEG or 2-DHG).

Example 3: Intrinsic carrier concentration in a 2DEG

Consider the density of states as shown in the Fig. 15.6 representing the two-dimensional density of states of electrons and holes in a quantum well.

(a) Assuming the well is undoped, obtain an expression for the Fermi level E_F at room temperature in terms of a, b and temperature T. Assume Boltzmann statistics to be valid. When is E_F exactly equal to the midgap energy, $\frac{(E_c+E_v)}{2}$?

(b) Obtain the expression for n_i, the intrinsic carrier concentration.

Hint: Start with the approximate expressions for the electron (n) and hole (p) concentrations in terms of $g_c(E)$ and $g_v(E)$ [subscripts c and v denote conduction and valence bands], and assume Boltzmann statistics of carriers:

$$f(E) = e^{\frac{E_F - E}{k_B T}}, \tag{15.168}$$

where k_B is Boltzmann's constant.

Solution

(a) The electron and hole concentrations are given by

$$n = \int_{E_c}^{\infty} g_c(E) f(E) dE, \tag{15.169}$$

$$p = \int_{-\infty}^{E_v} g_v(E)[1 - f(E)] dE. \tag{15.170}$$

Use the Boltzmann approximation, $f(E) = e^{\frac{(E_F - E)}{k_B T}}$ and $g_c(E) = a\theta(E - E_c)$ and $g_v(E) = b\theta(E_v - E)$ where $\theta(x) = 1$ for $x > 0$ and $\theta(x) = 0$ for $x \leq 0$ (Heaviside function). We get

$$n = \int_{E_c}^{\infty} a e^{\frac{(E_F - E)}{k_B T}} dE = a k_B T e^{\frac{E_F - E_c}{k_B T}}. \tag{15.171}$$

Similarly,

$$p = \int_{-\infty}^{E_v} b e^{\frac{(E - E_F)}{k_B T}} dE = b k_B T e^{\frac{E_v - E_F}{k_B T}}. \tag{15.172}$$

If the sample is intrinsic, then $n = p = n_i$. Therefore,

$$a k_B T e^{(\frac{E_F - E_c}{k_B T})} = b k_B T e^{(\frac{E_v - E_F}{k_B T})}, \tag{15.173}$$

from which we derive

$$E_F = \frac{E_v - E_c}{2} + \frac{k_B T}{2} ln(\frac{b}{a}). \tag{15.174}$$

Hence, $E_F = \frac{E_v + E_c}{2}$ whenever $a = b$.

(b) The intrinsic carrier concentration is given by $n_i = \sqrt{np}$. Hence, using Equations (15.171) and (15.172) above, we obtain

$$n_i = k_B T \sqrt{ab} e^{\frac{(E_c - E_v)}{2k_B T}} = k_B T \sqrt{ab} e^{-\frac{Eg}{2k_B T}}. \tag{15.175}$$

Example 4: Connection between 2D and 3D density of states

The density of states in the conduction band of a bulk sample is given by Equation (15.110).

If a 2D quantum well (of width W) is formed with infinite barriers on both sides, show that

$$g_{3D}(E_{c0} + E_n) = \frac{n}{W} g_{2D}(E_{c0} + E_n), \tag{15.176}$$

where $g_{2D}(E)$ is the two-dimensional density of states in each subband of the 2 DEG, given by Equation (15.127).

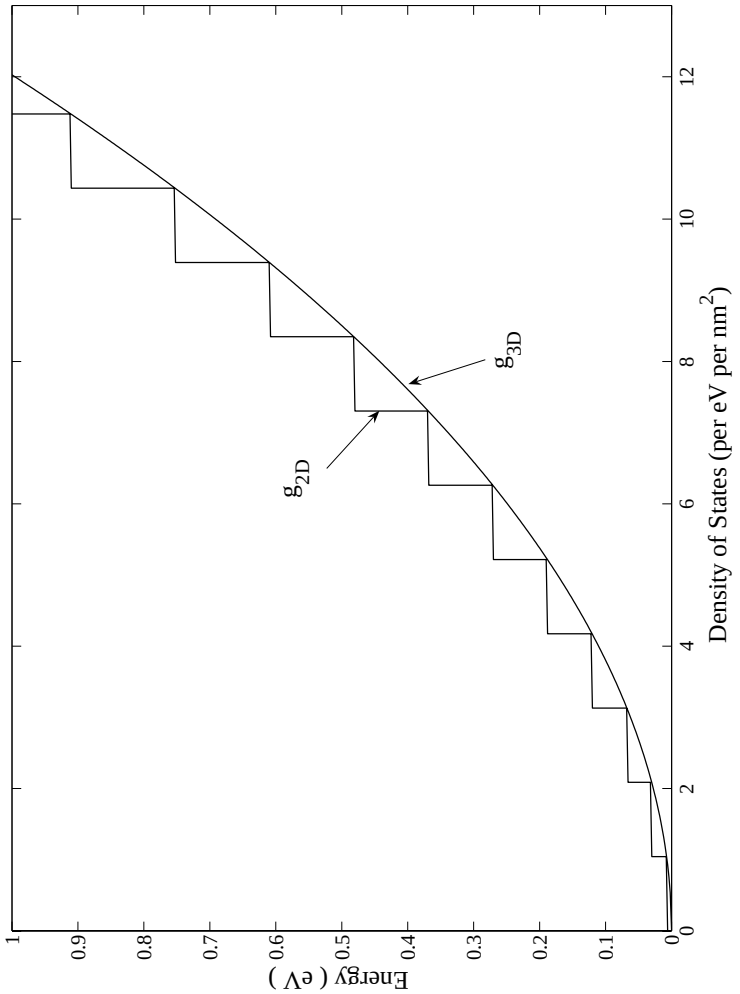


FIGURE 15.7
Illustration of the 2D density-of-states staircase touching the 3D density of states curve. For this illustration, the effective mass of electrons was assumed to be $m^* = 0.5m_o$ and the quantum well width was assumed equal to $100\mathring{A}$.

Solution

For a particle in a box with a constant potential energy E_{co}, the eigenenergies are given by

$$E_n = E_{co} + \frac{n^2\hbar^2\pi^2}{2m^*W^2}. \tag{15.177}$$

Hence, using Equation (15.110) we get

$$g_{3D}(E_n) = \frac{m^*}{\pi^2\hbar^3}\sqrt{2m^*(\frac{n^2\hbar^2\pi^2}{2m^*w^2})}, \tag{15.178}$$

i.e.,

$$g_{3D}(E_n) = \frac{m^*}{\pi^2\hbar^2}\frac{n}{w} = \frac{n}{w}g_{2D}(E_{co} + E_n). \tag{15.179}$$

A plot of $Wg_{3D}(E_n)$ and $g_{2D}(E_{co} + E_n)$ is shown in Fig. 15.7. This figure shows that the corners of the staircase representing the $\frac{m^*}{\pi\hbar^2}$ jump for each appearance of a new subband in the 2DEG touches the curve $Wg_{3D}(E_n)$. As the well width is increased, the energy levels for the particle-in-a-box are more closely spaced and the staircase becomes closer and closer to the Wg_{3D} curve, i.e.,

$$g_{3D}(E) = \lim_{W\to\infty}\frac{1}{W}g_{2D}(E). \tag{15.180}$$

15.6 The Rayleigh-Ritz variational procedure

The Rayleigh-Ritz variational principle is often used when the wavefunctions of the ground state and the first few excited states of a Hamiltonian cannot be found analytically, and when perturbation theory is too poor of an approximation to calculate the energies of the lower eigenstates of the Hamiltonian. In that case, physical arguments are invoked to *guess* an analytical form for the eigenstates of the Hamiltonian in terms of some variational parameters. The latter are then varied until the expectation value of the Hamiltonian (or the energy eigenvalue) is minimized. More precisely, the Rayleigh-Ritz variational procedure is based on the premise that if $|\phi>$ is a guess for the ground state wavefunction, then the actual ground state energy will satisfy the following inequality:

$$E_o \leq E_{min} = min(\frac{<\phi|H|\phi>}{<\phi|\phi>}), \tag{15.181}$$

where $|\phi>$ is the trial wavefunction which contains variational parameters. The denominator appearing on the right hand side makes sure that the trial wavefunction is normalized. The variational parameters are varied until the expression in the right hand side is minimized. This yields the best guess for $|\phi>$.

The variational method can also be applied to obtain the eigenfunctions of the higher excited states but, in that case, the trial function for the excited state must be selected such that it is orthogonal to the trial eigenfunctions selected to describe the wavefunctions of the lower energy eigenstates.

A more thorough discussion of the Rayleigh-Ritz variational principle can be found in some quantum mechanics textbooks. In this chapter, we illustrate this principle with a few sample problems, including an estimation of the ground state energy in a triangular well, which approximates the potential energy profile in a High Electron Mobility Transistor (HEMT) close to the heterointerface [see Fig. 12.6]. Since all spin field effect transistors that we discussed in this book (and a few others) employ HEMT structures, this result is important for the detailed analysis of these transistors.

Examples

Example 1: The harmonic oscillator

Suppose we start with the following trial wavefunction for the harmonic oscillator ground state:

$$\phi(x) = \frac{N}{x^2 + a^2}, \tag{15.182}$$

where a is a variational parameter and N is an extra factor needed to normalize the wavefunction.

This choice seems reasonable since the wavefunction is peaked at $x = 0$ and decays for large value of x. Moreover, this form of $\phi(x)$ allows an exact calculation of the functional E_{min} appearing in the inequality (15.181).

Now, apply the variational principle and find E_{min}. How does it compare with the true eigenvalue which is $\frac{\hbar\omega}{2}$, where ω is the angular frequency appearing in the expression for the potential energy $V(x) = \frac{m\omega^2 x^2}{2}$?

Solution

The normalization coefficient N is found to be

$$N = \sqrt{\frac{2a^3}{\pi}}, \tag{15.183}$$

and the energy E_{min} on the right hand side of inequality (15.181) is given by

$$E_{min} = \frac{h^2 + 2a^4 m^{*2}\omega^2}{4m^* a^2}. \tag{15.184}$$

The latter reaches a minimum for

$$a_{min} = \frac{1}{2^{1/4}}\sqrt{\frac{h}{m^*\omega}}. \tag{15.185}$$

The corresponding minimum energy E_{min} is

$$E_{min}(a_{min}) = \sqrt{2}\frac{\hbar\omega}{2}, \tag{15.186}$$

which is $\sqrt{2}$ times larger than the exact result $\hbar\omega/2$.

Example 2: The triangular well problem

The following two variational functions are used to describe the two lowest states in the two-dimensional electron gas formed at the interface of a HEMT. Close to the heterointerface, the conduction band in the semiconductor substrate is modeled as a triangular potential well [see Fig. 12.6]. The trial wavefunctions for the ground and first excited states are:

$$\xi_o(x) = (\frac{b_o^3}{2})^{\frac{1}{2}} x e^{\frac{-b_o x}{2}}, \tag{15.187}$$

$$\xi_1(x) = (\frac{3b_1^5}{2[b_o^2 + b_1^2 - b_o b_1]})^{\frac{1}{2}} x(1 - \frac{b_o + b_1}{6}x)e^{-\frac{b_1 x}{2}}, \tag{15.188}$$

where b_o and b_1 are two variational parameters, and $x = 0$ at the heterointerface.

Notice that both $\xi_o(x)$ and $\xi_1(x)$ are zero at $x = 0$, as they should be since neither wavefunction can penetrate the barrier at the heterointerface if it is very high. Since the potential energy increases linearly with distance into the substrate, the choice of exponentially decaying trial wavefunctions seems appropriate.

- Show that $\xi_o(x)$, $\xi_1(x)$ are an appropriate set of trial wavefunctions for the ground state and first excited states, i.e., $\xi_o(x)$ and $\xi_1(x)$ are both normalized, so that $\int_0^\infty |\xi_i(x)|^2 dx = 1$, and ξ_o and ξ_1 are orthogonal, meaning $\int_0^\infty \xi_o(x)\xi_1(x)dx = 0$.

Solution

Performing the integration over the probability in the ground state (magnitude of the square of the wavefunction), we get

$$\int_0^{+\infty} |\xi_o(x)|^2 dx = \int_0^{+\infty} (\frac{b_o^3}{2})x^2 e^{-b_o x} dx = \frac{b_o^3}{2}\frac{2}{b_o^3} = 1. \tag{15.189}$$

So ξ_o is indeed normalized.

Similarly,

$$\int_0^{+\infty} |\xi_1(x)|^2 dx = \frac{3b_1^5}{2(b_o^2 + b_1^2 - b_o b_1)}[I_2 - I_3 + I_4], \tag{15.190}$$

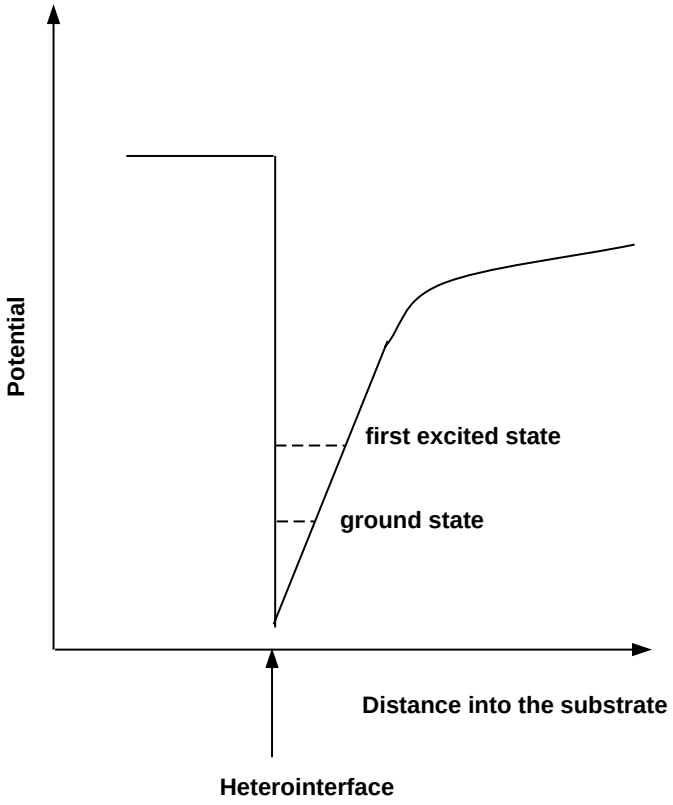


FIGURE 15.8
A triangular shape approximates the conduction band profile in the channel of a HEMT. The bound states and the wavefunctions of these states determine the channel carrier concentration and the threshold voltage.

where

$$I_2 = \int_0^{+\infty} x^2 e^{-b_1 x} dx, \tag{15.191}$$

$$I_3 = \frac{(b_o + b_1)}{3} \int_0^{+\infty} x^3 e^{-b_1 x} dx, \tag{15.192}$$

and

$$I_4 = (\frac{b_o + b_1}{6})^2 \int_0^{+\infty} x^4 e^{-b_1 x} dx. \tag{15.193}$$

We find

$$I_2 = \frac{2}{b_1^3}, \tag{15.194}$$

$$I_3 = \frac{2(b_o + b_1)}{b_1^4}, \tag{15.195}$$

and

$$I_4 = \frac{2}{3} \frac{(b_o + b_1)^2}{b_1^5}. \tag{15.196}$$

Hence

$$\int_0^{+\infty} |\xi_1(x)|^2 dx = \frac{3}{2} \frac{b_1^5}{[b_o^2 + b_1^2 - b_o b_1]} [\frac{2}{b_1^3} - \frac{2(b_o + b_1)}{b_1^4} + \frac{2(b_o + b_1)^2}{3b_1^5}]. \tag{15.197}$$

Simplifying

$$\int_0^{+\infty} |\xi_1(x)|^2 dx = \frac{3}{2} \frac{1}{[b_o^2 + b_1^2 - b_o b_1]} \frac{2}{3} [b_o^2 + b_1^2 - b_o b_1] = 1. \tag{15.198}$$

So $\xi_1(x)$ is also normalized.

To prove the orthogonality of ξ_o and ξ_1, we must show that the following integral is equal to zero

$$\int_0^{+\infty} \xi_o^*(x) \xi_1(x) dx = (\frac{b_o^3}{2})^{\frac{1}{2}} (\frac{3b_1^5}{2[b_o^2 + b_1^2 - b_o b_1]})^{\frac{1}{2}} (J_2 - J_3), \tag{15.199}$$

where

$$J_2 = \int_0^{+\infty} x^2 e^{-(\frac{b_o + b_1}{2})x} dx, \tag{15.200}$$

and

$$J_3 = (\frac{b_o + b_1}{6}) \int_0^{+\infty} x^3 e^{-(\frac{b_o + b_1}{2})x} dx. \tag{15.201}$$

We find

$$J_2 = J_3 = \frac{16}{(b_o + b_1)^3}. \tag{15.202}$$

Hence

$$\int_0^{+\infty} \xi_o(x) \xi_1(x) dx = 0, \tag{15.203}$$

and ξ_o, ξ_1 are indeed orthogonal.

- Using the results above, calculate an upper bound for the energy of the ground state in a triangular well in which the potential is assumed to be equal to infinity for $x < 0$ and for which the potential energy is given by

$$E_c(x) = \beta x, \tag{15.204}$$

for $x > 0$.

This problem can actually be solved exactly but requires the use of Airy functions which are rather cumbersome to deal with. The exact result for the ground state energy is given by [19]

$$E_0 = 1.857(\frac{\beta^2\hbar^2}{m^*})^{1/3}. \tag{15.205}$$

Compare your answer obtained using the trial wavefunction with the exact value above.

Solution

The expectation value E_{min} appearing in the inequality (15.181) is easily calculated using MATHEMATICA [20] or other software tools such as MATLAB that allow symbolic manipulation. The minimum of E_{min} occurs for

$$b_{o,min} = (\frac{12\beta m^*}{\hbar^2})^{1/3}, \tag{15.206}$$

leading to an upper bound on the ground state energy

$$E_{min}(b_{0,min}) = 1.966(\frac{\beta^2\hbar^2}{m^*})^{1/3}, \tag{15.207}$$

which is only about 6% larger than the exact result.

15.7 The transfer matrix formalism

The transfer matrix method is a powerful technique used to solve tunneling problems in quantum mechanics. It can calculate the transmission and reflection coefficients of an electron wave impinging on an arbitrary conduction band energy profile, as shown in Fig 15.9, provided there are no dissipative interactions such as phonon or electron-electron scattering that destroy the electron's phase memory. Non-phase-breaking scattering potentials (i.e., time-independent scattering potentials due to impurities, etc.) can be handled within this formalism since they simply add (or subtract) a spatially varying potential to the conduction band profile. The transfer matrix method is used

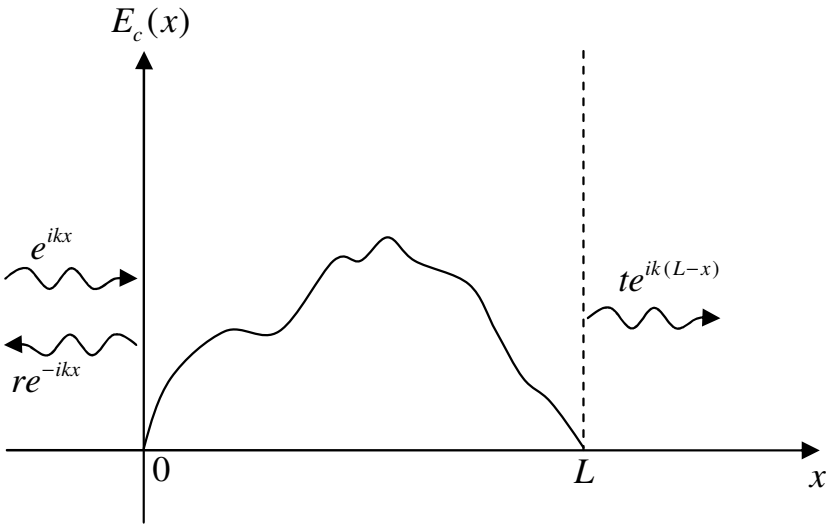


FIGURE 15.9
Arbitrary conduction band energy profile $E_c(x)$ in a region of length L (the device) sandwiched between two contacts; r and t are the reflection and transmission amplitudes, respectively, for an electron incident from the left contact.

to study quantum transport (including spin dependent transport through such devices as Spin Field Effect Transistors) and in modeling passive devices such as spin valves (see Chapter 11).

We wish to calculate the transmission and reflection coefficients of an electron wave incident on an arbitrary conduction band energy profile in a region of length L ("the device") sandwiched between a left (x $<$ 0) and a right (x $>$ L) contact, as illustrated in Fig. 15.9. The structure is shown under zero bias with similar contacts. Electrons are assumed to travel across the device without losing energy, i.e., the transport of carriers is assumed to be coherent throughout the structure. We proceed as follows.

15.7.1 Linearly independent solutions of the Schrödinger equation

We consider the one-dimensional time-independent Schrödinger equation assuming that only the potential energy $E_c(x)$ is varying in the x-direction, as shown in Fig. 15.9. If the effective mass of the electron is assumed to be independent of x, the Schrödinger equation becomes

$$-\frac{\hbar^2}{2m^*}\ddot{\phi}(x) + E_c(x)\phi(x) = E_p\phi(x), \tag{15.208}$$

where $\ddot{\phi}(x)$ stands for $\frac{d^2}{dx^2}\phi(x)$, the second derivative with respect to x, and E_p is the longitudinal kinetic energy, i.e., the kinetic energy in the x-direction, assumed to be the direction of current flow.

The general solution of this second order differential equation for $\phi(x)$ can be written as a linear combination of two linearly independent solutions [21]. Two solutions ϕ_1 and ϕ_2 of a differential equation are linearly independent if the equality $c_1\phi_1 + c_2\phi_2 = 0$ cannot be satisfied $\forall(x)$ for any choice of (c_1, c_2) except for $c_1 = c_2 = 0$. If non-zero solutions (c_1, c_2) exist, then ϕ_1 and ϕ_2 are said to be linearly dependent.

15.7.2 Concept of Wronskian

If ϕ_1 and ϕ_2 are linearly dependent, $c_1\phi_1 + c_2\phi_2 = 0$ and $c_1\dot{\phi}_1 + c_2\dot{\phi}_2 = 0$, where the dot stands for the first derivative with respect to x. Hence, in a matrix form, we have

$$\begin{bmatrix} \dot{\phi}_1 & \dot{\phi}_2 \\ \phi_1 & \phi_2 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}. \tag{15.209}$$

This means that the Wronskian ($W(x) = \dot{\phi}_1\phi_2 - \phi_1\dot{\phi}_2$), which is the determinant of the 2×2 matrix in the last equation, must be identically zero or otherwise only the trivial solution $c_1 = c_2 = 0$ in the last equation would be admissible. Stated otherwise, for two solutions to be linearly independent, their Wronskian must be non-zero for all x.

Even if $E_c(x)$ has finite discontinuities, $\dot{\phi}(x)$ exists throughout and, hence, $\phi(x)$ and $\dot{\phi}(x)$ are continuous. Since (ϕ_1, ϕ_2) satisfy the Schrödinger equation, we have

$$\ddot{\phi}_1\phi_2 - \ddot{\phi}_2\phi_1 = 0, \tag{15.210}$$

or

$$\frac{d}{dx}(\dot{\phi}_1\phi_2 - \dot{\phi}_2\phi_1) = \frac{dW}{dx} = 0. \tag{15.211}$$

Thus, $W(x)$ is a constant independent of x.

In summary, if the Wronskian $\mathrm{W(x)} = \dot{\phi}_1\phi_2 - \phi_1\dot{\phi}_2 = 0$, (ϕ_1, ϕ_2) are linearly dependent and, if $\mathrm{W(x)} \neq 0$, it is a constant and the two solutions are linearly independent.

If we can find these two linearly independent solutions, their linear combination is the most general solution to the Schrödinger equation,

$$\phi(x) = c_1\phi_1(x) + c_2\phi_2(x). \tag{15.212}$$

If we select $\phi_1(x)$ and $\phi_2(x)$ such that

$$\phi_1(0) = 0, \dot{\phi}_1(0) = 1, \tag{15.213}$$

and

$$\phi_2(0) = 1, \dot{\phi}_2(0) = 0, \tag{15.214}$$

they will be linearly independent since their Wronskian, which is independent of x, is equal to

$$W(x) = W(0) = \dot{\phi}_1(0)\phi_2(0) - \phi_1(0)\dot{\phi}_2(0) = 1. \tag{15.215}$$

15.7.3 Concept of transfer matrix

Since ϕ_1 and ϕ_2 satisfy the conditions (15.213-15.214), we have

$$c_1 = \dot{\phi}(0), \tag{15.216}$$

and

$$c_2 = \phi(0). \tag{15.217}$$

Hence, from Equation (15.212) and its first derivative with respect to x, we get

$$\phi(L) = \dot{\phi}(0)\phi_1(L) + \phi(0)\phi_2(L), \tag{15.218}$$

and

$$\dot{\phi}(L) = \dot{\phi}(0)\dot{\phi}_1(L) + \phi(0)\dot{\phi}_2(L), \tag{15.219}$$

which can be rewritten in a matrix form:

$$\begin{bmatrix} \dot{\phi}(L) \\ \phi(L) \end{bmatrix} = \begin{bmatrix} \dot{\phi}_1(L) & \dot{\phi}_2(L) \\ \phi_1(L) & \phi_2(L) \end{bmatrix} \begin{bmatrix} \dot{\phi}(0) \\ \phi(0) \end{bmatrix}. \tag{15.220}$$

The matrix appearing on the right hand side is called the transfer matrix because it relates (or connects) the column vector $(\dot{\phi}(x), \phi(x))^\dagger$ (where the $\dagger$ stands for the transpose operation) at location $x = L$ to its value at location $x = 0$.

15.7.4 Cascading rule for transfer matrices

An arbitrary spatially varying potential energy profile $E_c(x)$ can always be decomposed into a series of steps where the potential energy in each step is approximated by a spatially invariant constant which is its average value over that interval. The accuracy of this approximation increases with decreasing interval size. Since the potential within each section is constant, the transfer matrix of each small section can be derived exactly. Once the individual transfer matrix for each small segment is known, the overall transfer matrix W_{TOT} needed to relate the wavefunction on the right to the wavefunction on the left (i.e., describe the tunneling process of a particle incident from the left contact) is the product of the individual transfer matrices associated with each small segment, i.e.,

$$W_{TOT} = W_N W_{N-1} ... W_2 W_1, \tag{15.221}$$

where W_i is the transfer matrix associated with the i^{th} segment counted from the left contact.

Since matrices do not commute in general, it makes a difference whether we start cascading from the left or the right. A little introspection should convince the reader that the cascading should proceed from right to left as shown in the equation above. The tunneling problem for a particle incident from the right contact can be treated using the same formalism by flipping the structure around, which is equivalent to taking the product of the different segments in the reverse order.

From now on, we use the notation W for the overall transfer matrix W_{TOT} and label its elements W_{ij} with $(i, j) = 1$ or 2. As shown in Fig. 15.9, the boundary conditions for a particle incident from the left contact are given by

$$\phi(x) = e^{ikx} + re^{-ikx}, \tag{15.222}$$

and

$$\phi(x) = te^{ik(L-x)}, \tag{15.223}$$

where we have assumed that the contacts are identical and that there is no bias across the device (the case of non-zero bias is an easy extension).

From Equation (15.220), we get the two following equations for the reflection (r) and transmission (t) *amplitudes*

$$\begin{pmatrix} ikt \\ t \end{pmatrix} = \begin{pmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{pmatrix} \begin{pmatrix} ik(1-r) \\ 1+r \end{pmatrix}. \tag{15.224}$$

These two equations can be solved explicitly leading to

$$t = \frac{2ik}{(k^2 W_{21} - W_{12}) + ik\,(W_{11} + W_{22})}, \tag{15.225}$$

and

$$r = \frac{\left(k^2 W_{21} + W_{12}\right) + ik\,(W_{11} - W_{22})}{(k^2 W_{21} - W_{12}) + ik\,(W_{11} + W_{22})}. \tag{15.226}$$

Exercises:

- Each of the transfer matrices appearing in the product is unimodular, i.e., their Wronskian is unity. Prove that it is also true for the overall transfer matrix W_{TOT}.
- Derive the expressions of the transmission and reflection amplitudes given above by solving the system of two equations with two unknowns (r, t) in equation (15.224).
- Prove that the sum of the reflection coefficient, $|r|^2$, and transmission coefficient, $|t|^2$, is unity since the incident particle must be either reflected or transmitted.

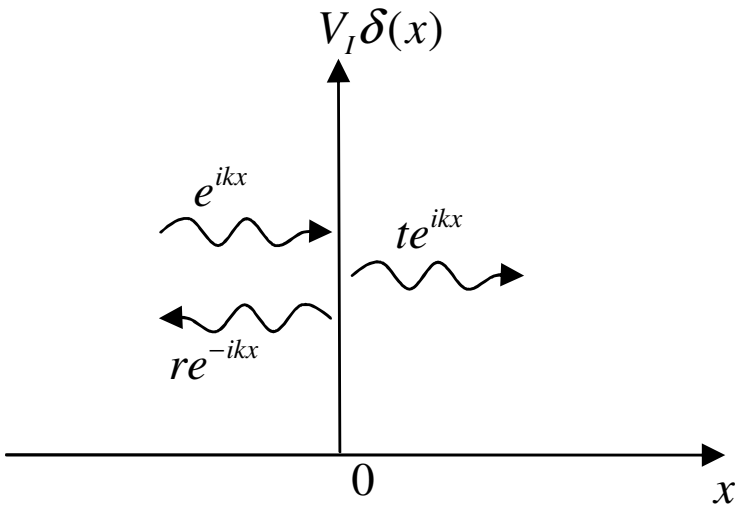


FIGURE 15.10
Scattering problem for an electron incident from the left on a one-dimensional delta-scatterer.

Next, we calculate explicitly the transfer matrix and transmission coefficient for a specific problem.

Example

Tunneling through a one-dimensional delta scatterer

Consider the one-dimensional scattering problem shown in the figure below where the scattering potential is modeled as a one-dimensional delta scatter, i.e., $V(x) = V_I\delta(x)$, located at $x = 0$. The parameter V_I is the strength of the delta-scatterer and has units S.I. units of Joules times meter. It is often given in the literature in units of eV$\mathring{A}$.

Starting with the Schrödinger equation

$$-\frac{\hbar^2}{2m^*}\ddot{\phi}(x) + V_I\delta(x)\phi(x) = E_p\phi(x), \tag{15.227}$$

and integrating it on both sides from 0_- to 0_+, we get

$$-\frac{\hbar^2}{2m^*}\left[\dot{\phi}(0_+) - \dot{\phi}(0_-)\right] + V_I\phi(0) = 0, \tag{15.228}$$

which leads to

$$\dot{\phi}(0_+) = \dot{\phi}(0_-) + \frac{2m^* V_I}{\hbar^2}\phi(0_+). \tag{15.229}$$

In addition, the wavefunction is assumed to be continuous across the delta scatterer, i.e.,

$$\phi(0_+) = \phi(0_-). \tag{15.230}$$

Regrouping the last two equations in a matrix form, we get

$$\begin{bmatrix} \dot{\phi}(0_+) \\ \phi(0_+) \end{bmatrix} = \begin{bmatrix} 1 & \frac{2m^* V_I}{\hbar^2} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \dot{\phi}(0_-) \\ \phi(0_-) \end{bmatrix}. \tag{15.231}$$

So, the transfer matrix across a delta scatterer is given by

$$W = \begin{bmatrix} 1 & 2k_\delta \\ 0 & 1 \end{bmatrix}, \tag{15.232}$$

where $k_\delta = \frac{m^* V_I}{\hbar^2}$.

Exercise: Find the reflection and transmission amplitudes for an electron incident from the left on the delta scatterer shown in Fig. 15.10 and prove that $|r|^2 + |t|^2 = 1$.

Solution

Equation (15.230) leads to

$$1 + r = t, \tag{15.233}$$

whereas Eq.(15.229) leads to

$$ikt - ik(1 - r) = 2k_\delta t. \tag{15.234}$$

Solving these last two equations for the two unknowns t and r, we get

$$t = \frac{ik}{ik - k_\delta}, \tag{15.235}$$

and

$$r = \frac{k_\delta}{ik - k_\delta}. \tag{15.236}$$

Using the last two results, it is readily shown that

$$|r|^2 + |r|^2 = 1. \tag{15.237}$$

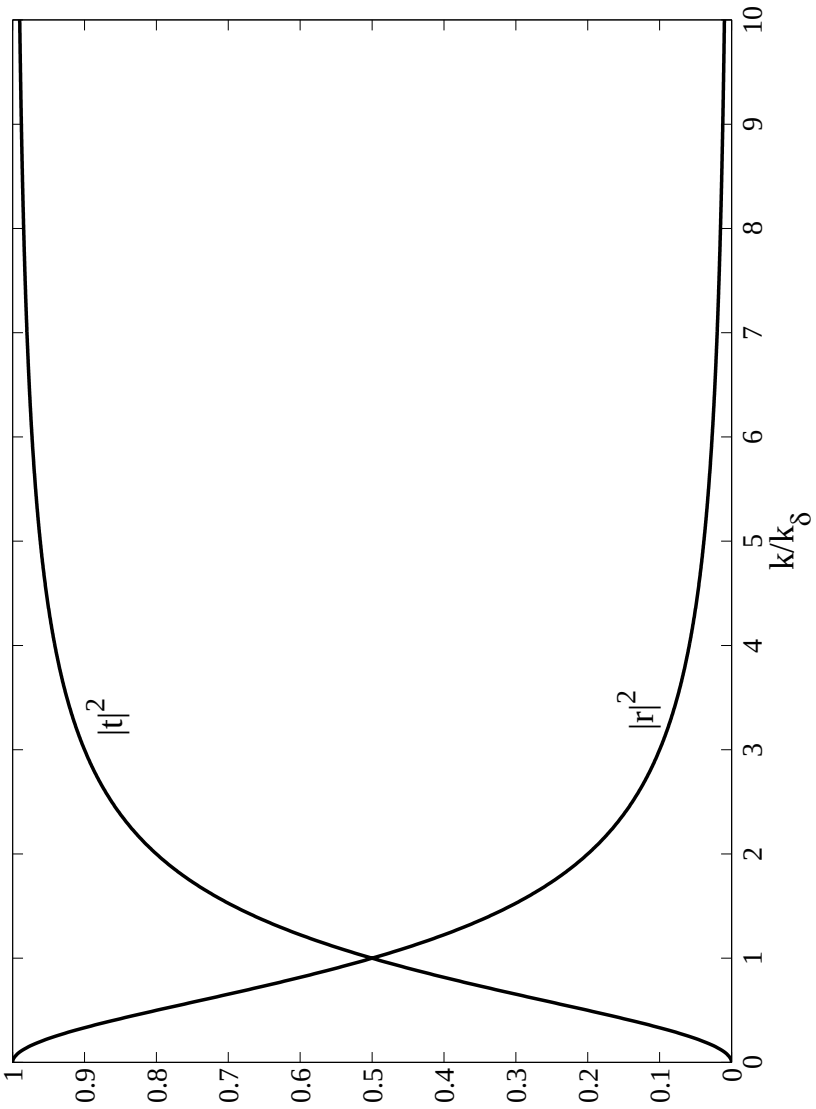


FIGURE 15.11
Plot of the transmission probability $|t|^2$ and reflection probability $|r|^2$ across a delta-scatterer as a function of the reduced eigenvector k/k_δ. $|t|^2 = |r|^2 = 0.5$ when $k/k_\delta = 1$.

15.8 Problems

- **Problem 15.1**

  Calculate the right hand side of the generalized Heisenberg inequality (15.54) for the case where the two operators C and D are given by

$$C = \sigma_+, D = \sigma_-, \tag{15.238}$$

  where $\sigma_+ = \sigma_x + i\sigma_y$ and $\sigma_- = \sigma_x - i\sigma_y$.

- Prove that for the case where the operators C and D are equal to x and p_x, the inequality becomes an equality sign for a one-dimensional harmonic oscillator in its ground state.

- Repeat the previous problem for an harmonic oscillator in its first excited state. What is the value of $\Delta x \Delta p_x$ in this case?

Solution

For the ground state and first excited states,

$$< x >= 0, \tag{15.239}$$

and

$$< p_x >= 0. \tag{15.240}$$

In the ground state,

$$< x^2 >= \frac{\hbar}{2m\omega}, \tag{15.241}$$

$$< {p_x}^2 >= \frac{\hbar m\omega}{2}, \tag{15.242}$$

and

$$\Delta x \Delta p_x = \frac{\hbar}{2}. \tag{15.243}$$

In the first excited state,

$$< x^2 >= \frac{3\hbar}{2m\omega}, \tag{15.244}$$

$$< {p_x}^2 >= \frac{3\hbar m\omega}{2}, \tag{15.245}$$

and

$$\Delta x \Delta p_x = \frac{3}{2}\hbar. \tag{15.246}$$

15.9 References

[1] L. I. Schiff, *Quantum Mechanics*, (McGraw-Hill, New York, 1955).

[2] J. J. Sakurai, *Modern Quantum Mechanics* (Addison Wesley, Redwood City, CA, 1985).

[3] D. J. Griffiths, *Introduction to Quantum Mechanics*, (Prentice Hall, Englewood Cliffs, New Jersey, 1995).

[4] H. Kroemer, *Quantum Mechanics for Engineering, Materials Science and Applied Physics*, (Prentice Hall, Englewood Cliffs, New Jersey, 1994).

[5] G. Gamow and J. M. Cleveland, *Physics*, (Prentice Hall, Englewood Cliffs, NJ, 1969).

[6] N. Bohr, "The quantum postulate and the recent development of atomic theory", *Naturwissenschaften*, **16**, 245 (1928).

[7] P. Ghose, D. Home and G. S. Agarwal, "An experiment to throw more light on light", Phys. Lett. A, **153**, 403 (1991).

[8] J. C. Bose, in: *Collected Physical Papers*, (Longman and Green, London, 1927) pp. 44-49.

[9] P. Grangier, G. Roger and A. Aspect, "Experimental evidence for a photon anti-correlation effect on a beam splitter: A new light on single photon interferences", Europhys. Lett., **1**, 173 (1986).

[10] Y. Mizobuchi and Y. Ohtaké, "An 'experiment to throw more light on light' ", Phys. Lett. A, **168**, 1 (1992).

[11] S. Rangwala and S. M. Roy, "Wave behaviour and non-complementary particle behaviour in the same experiment", Phys. Lett. A, **190**, 1 (1994).

[12] P. A. M. Dirac, *Principles of Quantum Mechanics*, 4th edition, (Oxford University Press, Oxford, 1958), p. 8.

[13] B-G Englert, "Fringe visibility and which way information: An inequality", Phys. Rev. Lett., **77**, 2154 (1996).

[14] W. K. Wootters and W. K. Zurek, "Complementarity in the double slit experiment: Quantum non-separability and a quantitative statement of Bohr's principle", Phys. Rev. D, **19**, 473 (1979).

[15] D. M. Greenberger and A. Yasin, "Simultaneous wave and particle knowledge in a neutron interferometer", Phys. Lett. A, **128**, 391 (1988).

[16] L. Mandel, "Coherence and indistinguishability", Opt. Lett., **16**, 1882 (1991).

[17] M.A. Nielsen and I.L. Chuang, *Quantum Computation and Quantum Information*, (Cambridge University Press, 2000).

[18] C. Cohen-Tannoudji, B. Diu and F. Laloe, *Quantum Mechanics*, (Hermann, Paris, France, 1977).

[19] D. ter Haar, *Problems in Quantum Mechanics*, (Pion, 1975).

[20] http://www.wolfram.com

[21] E. Kreyszig, *Advanced Engineering Mathematics*, (John Wiley & Sons, New York, 2006).

Nanotechnology

INTRODUCTION TO Spintronics

Using spin to replace or augment the role of charge in signal processing devices, computing systems and circuits may improve speed, power consumption, and device density in some cases—making the study of spin one of the fastest-growing areas in micro- and nano-electronics. With most of the literature on the subject still highly advanced and heavily theoretical, the demand for a practical introduction to the concepts relating to spin has only now been filled.

Introduction to Spintronics is an accessible, organized, and progressive presentation of the quantum mechanical concept of spin. The authors build a foundation of principles and equations underlying the physics, transport, and dynamics of spin in solid state systems. They explain the use of spin for encoding qubits in quantum logic processors and how spin-orbit interaction forms the basis for certain spin-based devices such as spintronic field effect transistors. They also discuss the effects of magnetic fields on spin-based device performance, highlighting effects such as giant magnetoresistance, the subject of the 2007 Nobel Prize in physics.

Features

- Provides the first comprehensive introduction to spintronics, the fastest growing area of electronics
- Compares the advantages of spin versus charge for electrical devices such as transistors, diodes, and solar cells
- Assumes little prior background, only basic device physics and quantum mechanics
- Contains numerous examples and problems for practical illustration of the wide range of topics presented

Coverage spans active hybrid spintronic devices, monolithic spintronic devices, passive spintronic devices, and devices based on the giant magnetoresistance effect. The final chapters introduce the burgeoning field of spin-based reversible logic gates, spintronic embodiments of quantum computers, and other topics in quantum mechanics that have applications in spintronics. **Introduction to Spintronics** provides the knowledge and understanding of the field needed to conduct independent research in spintronics.

CRC Press
Taylor & Francis Group
an informa business
www.taylorandfrancisgroup.com

6000 Broken Sound Parkway, NW
Suite 300, Boca Raton, FL 33487
270 Madison Avenue
New York, NY 10016
2 Park Square, Milton Park
Abingdon, Oxon OX14 4RN, UK

3133
ISBN 0-8493-3133-1
90000
9 780849 331336
www.crcpress.com